Title

Experimental considerations for precise RNA-mediated insertion of transgenes

ISBN

9780443297243

Authors

Palm, Sarah M

Van Treeck, Briana

Collins, Kathleen

Publication Date

2024

DOI

10.1016/bs.mie.2024.07.007

HHS Public Access

Author manuscript

Methods Enzymol. Author manuscript; available in PMC 2025 October 18.

Published in final edited form as:

Methods Enzymol. 2024 ; 705: 1–24. doi:10.1016/bs.mie.2024.07.007.

Experimental considerations for precise RNA-mediated insertion of transgenes

Sarah M. Palm[1],

Briana Van Treeck[1],

Kathleen Collins[*]

Department of Molecular and Cell Biology, University of California at Berkeley, Berkeley, CA, United States

Abstract

Precise RNA-mediated insertion of transgenes (PRINT) is a pioneering method for site-specific, safe-harbor transgene supplementation of the human genome that harnesses a eukaryotic retroelement protein and relies solely on the delivery of RNA. Here we outline important considerations in the design of the two required RNAs, details for the production and transfection of these RNAs to cells, and read-outs for successful transgene addition. Throughout, tips and key concepts are laid out to enable general use of this method.

1. Introduction

Precise RNA-mediated INsertion of Transgenes (PRINT) is a complementary approach to existing gene editing methods that fulfills a long-standing challenge: the safe insertion of autonomously expressed, full-length genes into the human genome (Zhang et al., 2024). Engineered nucleases, for example Cas9 ribonucleoproteins (RNPs), excel in gene disruption (Li et al., 2023). For transgene insertion, however, co-delivery of DNA is required, and the insertion mechanism relies on homology-directed repair, which is restricted to genome-replication phases of the cell cycle and even then has low efficiency relative to gene disruption (Yang et al., 2020). Viral delivery of transgenes is restricted by packaging limitations and adaptive immune responses. Also, transgenes remain episomal or integrate with sites of insertion that are random enough to make oncogenic mutagenesis a major risk (Hacein-Bey-Abina et al., 2008). PRINT features many desired characteristics for a gene addition platform (Zhang et al., 2024). First, stable genome insertion of transgenes is exquisitely site-specific, limiting off-target effects. Second, there is no requirement for the introduction of DNA; in fact, there is no extrachromosomal DNA at any step of gene insertion. Finally, there is strict selectivity for RNA transcripts used to template gene insertion, such that only the transgene of interest will be inserted, not pseudogenes made from endogenous cellular RNAs.

[*]Corresponding author. kcollins@berkeley.edu.

[1]These authors contributed equally.

Here we detail the experimental reagents and steps for successful implementation of PRINT. In brief, two RNAs are introduced to the cell: (1) an mRNA encoding an R2 retroelement protein (R2p) and (2) the template RNA that will be reverse transcribed into the genome (Fig. 1A). The mRNA is designed to be efficiently translated within the cell and produces the R2p within a few hours. R2p binds specifically to the 3′ end of the template RNA, and the RNP finds its target site within the highly repetitive rRNA genes (rDNA) in the nucleolus (Fig. 1B, Eickbush & Eickbush, 2015; Zhang et al., 2024). Once bound to its target site, R2p initiates target-primed reverse transcription (TPRT) (Fig. 1C, Luan et al., 1993). The target site is nicked by the enzyme, and the nicked strand 3′ end is used as the primer for first-strand cDNA copying from the template RNA. Consequently, cDNA synthesis occurs directly into the genome. Insertion is completed by second-strand cleavage and second-strand synthesis.

2. The right transposable element for the task

When thinking about the ability to insert genes into the genome, transposable elements (TEs) immediately come to mind. TEs are ubiquitous and diverse, and they are capable of propagating their own sequences to new locations throughout the genome (Bourque et al., 2018; Wells & Feschotte, 2020). TEs come in two flavors: DNA transposons and RNA transposons (retrotransposons). DNA transposons have been adapted to varying levels of success for non-native gene integration by a “cut and paste” mechanism (Sandoval-Villegas et al., 2021), in which the transposase enzyme binds the target site and makes concerted strand cleavages and phosphodiester bond exchanges (Fig. 2A). Drawbacks to using DNA transposons for gene delivery applications include the need to supply a foreign DNA substrate for integration, which is challenging both in delivery and in stimulation of genomic stress and pathogen response pathways. Retrotransposons propagate using RNA as an intermediate for “copy and paste” mechanisms. This difference in the type of templating nucleic acid gives retrotransposon-based gene delivery applications an appeal of potentially supplying the cell with the retroelement protein as well as a template RNA without the need to directly introduce foreign DNA into cells.

Eukaryotic retrotransposon classes include both long terminal repeat (LTR) and non-LTR branches. LTR retrotransposons and retroviruses first convert a template RNA into double-stranded DNA, then use an integrase enzyme to introduce DNA into a target site (“copy and paste”; Fig. 2B). Non-LTR retrotransposons differ from LTR retrotransposons and retroviruses in the synthesis of cDNA directly into the genome, with a concerted “nick and copy” mechanism of first-strand synthesis and then a “paste” step to form the second insertion junction (“nick-copy-paste”; Fig. 2C). The non-LTR category is unique in use of the target site for priming cDNA synthesis (Luan et al., 1993), a specificity designated as TPRT, which makes the entire insertion process free of an extrachromosomal DNA intermediate (Fig. 1C). The intramolecular coordination of endonuclease and reverse transcriptase activities in a single retrotransposon protein greatly reduces the existence of unprotected nicks or breaks (Zhang et al., 2024).

PRINT exploits the non-LTR retrotransposon R2, which we selected due to its exquisite target-site specificity for rDNA (Fig. 3, top; Yang & Eickbush, 1998; Lee et al., 2024).

Native R2 retrotransposons have used this highly conserved and multicopy target site as a safe landing pad for hundreds of thousands of years (Burke et al., 1998; Kojima & Fujiwara, 2005). In fact, some species can live and propagate without any apparent disadvantage from having a substantial percentage of their rDNA containing retroelement insertions (Jakubczak et al., 1991; Eickbush & Eickbush, 2003). While mammals lack R2, they maintain the highly conserved target site, which is identical between humans and the avian species whose R2ps are used in PRINT (Table 1). The lack of a mammalian R2 is optimal because an introduced R2 system would not crosstalk with endogenous R2 components, and an endogenous R2 could not compound the inactivation of functional rDNA units. Because the target site is highly conserved, it was reasonable to expect that R2ps from other eukaryotic species would hone in on this target site and use it productively for new DNA insertion (Fig. 3, bottom).

R2 has a broad phylogenetic distribution for two highly represented clades (A and D) and more restricted distribution for additional possible clades (B and C) (Luchetti & Mantovani, 2013; Kojima et al., 2016). It is present in insects, crustaceans, and non-mammalian vertebrates among other animals (Kojima et al., 2016). We screened predicted R2ps, mostly from arthropods, fish, and birds, for site-specific endonuclease and template-selective reverse transcriptase activities using recombinant proteins produced in human cells. Comparisons led us to the selection of the avian R2ps from *Taeniopygia guttata* (zebra finch, TaGu) and *Zonotrichia albicollis* (white-throated sparrow, ZoAl) for additional investigation. These avian A-clade R2ps demonstrate two safeguards for specificity. First, they are highly selective for using RNAs with avian R2 3′ untranslated regions (UTRs), which are conserved in sequence and structure, as TPRT templates. Second, they require template 3′-end base-pairing with the nicked target-site primer to initiate cDNA synthesis. Dual specificity of avian R2ps for both the template RNA and target site is ideally suited to PRINT.

By separating the R2 open reading frame (ORF) from the UTRs of an R2 RNA into two RNAs with divided function (mRNA and template RNA), we ensured that the R2p will **not** insert its own ORF sequence into the genome. Instead, R2p will use the template RNA sequence for insertion synthesis, since it has an R2 3′ UTR.

3. R2p mRNA design

The first component of PRINT is an mRNA that upon transfection into cells will be translated into R2p. R2ps are translated from one ORF, with all their biochemical activities coordinated as domains of a single polypeptide. This allows the use of just one mRNA to provide the required insertion machinery not already present in a cell (Fig. 4A). The ORFs for avian R2ps ZoAl and TaGu were codon-optimized for expression in mammalian cells and flanked by the 5′ and 3′ UTRs from the BioNTech COVID-19 mRNA vaccine (World Health Organization, 2020), all cloned into a plasmid with T7 RNA Polymerase (RNAP) promoter. We recommend the use of a modified uridine nucleotide such as pseudouridine (Ψ) or 1-methyl-Ψ (m1Ψ) for mRNA in vitro transcription (IVT) as this would decrease the innate immune response to RNA delivery (Karikó et al., 2008; To & Cho, 2021), but unmodified uridine can also be used for efficient PRINT (Zhang et al., 2024). To prepare the plasmid template for IVT, it is first linearized to completion by restriction enzyme digestion.

We prefer to design a unique *Bbs*I site immediately following the intended template RNA sequence; however, any Type IIS restriction enzyme that cleaves outside of its recognition sequence can be used to remove the restriction site from the IVT mRNA such that a polyadenosine (polyA) tract is 3′-terminal on the transcript (Fig. 4A, inset). Whether using *Bbs*I or a different Type IIS restriction enzyme, it is essential to verify that your chosen restriction site is absent in the T7 RNAP promoter, mRNA UTRs, and ORF, or otherwise removed via synonymous mutation.

Variations of the avian R2ps have been created to modulate insertion efficiency and serve as experimental controls. A double side-chain substitution within the active site of the reverse transcriptase (RT) domain creates an RT-dead (RTD) protein that is a critical negative control for PRINT experiments (Fig. 4A). In addition, a double side-chain substitution within the endonuclease (EN) domain yields an EN-dead (END) protein. We also created EN-tuned (ENT) R2p variants to reduce target-site nicking and insertion efficiency (Fig. 4A). In cells with high PRINT efficiency, ENT R2p limits insertion copy number and thereby improves long-term stability and expression of transgenes (Zhang et al., 2024). In the human telomerase reverse transcriptase-immortalized retinal pigment epithelial (hTERT RPE-1) cells used for PRINT optimization, a copy number of ~5 insertions can be maintained long-term (Zhang et al., 2024), but this is likely to be dependent on cell type, proliferation rate, and other variables and warrants additional evaluation. As a recommendation, we currently suggest the use of ENT R2p for long-term transgene expression.

In the original PRINT mRNA transcription construct, the BioNTech 3′ UTR is fused to its DNA-templated, linker-containing polyA tail ($A_{30}L_{10}A_{70}$) (World Health Organization, 2020). This homopolymer region of the mRNA IVT plasmid is unstable with propagation in *Escherichia coli* (Trepotec et al., 2019). We advise that any subcloning or amplification of plasmids with this or another DNA-encoded polyA tail requires careful screening to ensure the integrity of the transcribed mRNA polyA tail. In our experience, the use of Stable Cells from New England Biolabs (NEB) and culture growth at 30 °C limits, but does not eradicate, A-tract loss. In addition to Sanger and full-plasmid sequencing, we recommend performing restriction enzyme digests to confirm retention of the complete polyA sequence in the mRNA transcription plasmid prior to IVT. Conveniently, the *Bbs*I site at the end of the IVT template mRNA sequence, the *Nde*I site located within the L_{10} linker of the polyA tail, and the *Nhe*I site located between the R2 3′ UTR and the polyA tail can be used to verify homopolymer integrity through single- and double-digestion checks (Fig. 4B). After integrity of the mRNA expression plasmid is verified, mRNA production via IVT can proceed (see below).

4. Template RNA design

PRINT template RNAs generally consist of optimized 5′ and 3′ modules flanking an expression cassette containing an RNAP II promoter, transgene of choice, and polyA addition signal sequence (PA) (Fig. 5). Like native R2 insertions (Eickbush & Eickbush, 2015), transgene insertions can be 5′-truncated (Zhang et al., 2024). Until truncated

insertions can be suppressed, the polarity of first-strand synthesis from 3′ to 5′ on template RNA should be considered as part of transgene template design.

The template RNA 3′ module is crucial to insertion because it contains critical parts of the R2 3′ UTR that R2p binds and subsequently uses to initiate transgene cDNA synthesis. We previously performed a cross-species TPRT comparison of R2ps and 3′ UTRs that revealed the avian R2ps' striking template preference for 3′ UTRs from R2 elements found in closely related avian species (Zhang et al., 2024). Of the tested 3′ module 3′ UTRs, the relatively compact 3′ UTR of the medium ground finch *Geospiza fortis* (GeFo) was optimal for PRINT with ZoAl and TaGu R2ps. In fact, the 3′-terminal 98 nucleotides (nt) of GeFo 3′ UTR is sufficient (Palm et al., 2024). Avian R2ps require 3′ UTR extension by about 4 nt of sequence immediately downstream of the R2 insertion site (R4) to base-pair with the TPRT primer and optimally an A-tract of about 22 nt (A22) (Fig. 5; Zhang et al., 2024).

In contrast to the protein-interacting 3′ module, the primary function of the 5′ module appears to be protection of the template RNA from degradation by 5′–3′ exonucleases. The template RNA 5′ modules optimal for PRINT, especially when using uridine rather than Ψ for template RNA synthesis, fold as a hepatitis delta virus (HDV)-like ribozyme (Rz) (Palm, Horton, Zhang, & Collins, 2024). HDV-like Rzs self-cleave to yield a 5′ OH shielded from nuclease access by a complex double-pseudoknot tertiary structure, likely contributing to their remarkably long half-life in human cells (Ferré-D′Amaré et al., 1998; Lévesque et al., 2002). The prediction of an HDV-like Rz 5′-terminal RNA structure is generally conserved across R2 RNAs from diverse species, potentially reflecting the need for RNA stabilization by a structure to substitute for the missing methyl-guanosine cap not present on the native RNAP I transcript (Eickbush et al., 2013). In native R2 context, Rz self-cleavage also would release the R2 RNA 5′ end from the rRNA precursor in which it is embedded.

In the context of PRINT, transgene insertion efficiency is increased by inclusion of an HDV-like Rz with high thermodynamic stability, such as the Rz from the flour beetle *Tribolium castaneum* B-clade R2 element (TriCasA) or a native HDV Rz, at the template RNA 5′ end (Zhang et al., 2024; Palm et al., 2024). Additionally, 5′ module sequence specifies the dominant mechanism of transgene 5′ junction formation with rDNA. Template RNA 5′ modules that include 28 nt of rRNA from the region immediately upstream of the R2 insertion site (R28) as their self-cleaved 5′-terminus generate precise 5′ junctions formed by annealing of upstream rDNA and the transgene cDNA 3′ end (Zhang et al., 2024; Palm et al., 2024). A 5′ module lacking upstream rDNA sequence instead generates imprecise 5′ junctions formed by end-joining (Palm et al., 2024). While this distinction in 5′ junction formation is not a major determinant of PRINT efficiency using the protocol specifications below, it will gain importance if DNA repair by end-joining can be suppressed to eliminate stable insertion of transgenes with a 5′ truncation.

In our prior works, several additional key template RNA optimizations were identified that substantially improve PRINT efficiency. Critically, complete replacement of uridine with structure-stabilizing uridine analogs such as m1Ψ or optimally Ψ for template RNA synthesis improves transgene insertion efficiency and the percentage of full-length transgenes by several-fold (Zhang et al., 2024; Palm et al., 2024). Although Ψ modification

inhibits the self-cleavage of some HDV-like Rzs, the Rz-independent stabilization that it provides confers relatively high PRINT efficiency even using template RNAs with 5′ modules that lack designed structure formation (Palm et al., 2024). This apparent stabilizing effect is consistent with the finding that Ψ modification extends the half-life of transfected mRNA (Karikó et al., 2008).

Because the optimal PRINT template RNA 5′ module includes an Rz that self-cleaves to gain its final fold, the 5′ module Rz is preceded by a 5′ leader sequence in the "pre-cleaved" IVT RNA product. Careful consideration should be given to the leader sequence: for example, some leader sequences could impede Rz self-cleavage by favoring the formation of alternative RNA structures (Chadalavada et al., 2000). For native or designed Rzs that self-cleave at the R28 position upstream of the insertion site, we generally include an 8 nt leader sequence corresponding to the rDNA region upstream of R28, designated as L8. If the template RNA 5′ module Rz does not efficiently self-cleave, L8 will extend R28 to generate a longer rRNA sequence and thereby a longer length of 3′ cDNA capable of annealing to the upstream target site rDNA with the same outcome of precise transgene insertion (Palm et al., 2024). Following self-cleavage during IVT, the leader sequence is removed by RNA purification (see below).

With any 5′ module, it may be beneficial to treat template RNAs with phosphatase following IVT. Template RNAs without a 5′ Rz have a 5′ triphosphate that can stimulate the innate immune response, for example as a substrate for RIG-I (Pichlmair et al., 2006). Even template RNAs with a 5′ Rz may benefit from phosphatase treatment if self-cleavage is incomplete. However, phosphatase treatment does increase the potential for hydrolytic cleavage of template RNA because the reaction requires RNA incubation with divalent ion (Chheda et al., 2024). Importantly, phosphatase treatment is likely less necessary when template RNA is synthesized with Ψ, because RIG-I activation is significantly attenuated by replacement of uridine with certain uridine analogs (Karikó et al., 2008).

Other notable influences on the template RNA contribution to PRINT efficiency include the sequence of the transgene RNAP II promoter. Use of a modified cytomegalovirus (CMV) promoter rather than the chicken beta actin hybrid intron (CBh) promoter, for example, yielded a remarkably large improvement in PRINT efficiency (Zhang et al., 2024). The molecular mechanism(s) conferring this difference remain to be completely defined and exploited in future template RNA designs. Finally, the percentage of transgenes with 5′ truncations appears to increase proportionally with template RNA length, meaning that longer template RNAs are likely to have a relatively lower number of complete transgene insertions. Ongoing efforts aim to mitigate this penalty by approaches including improved template RNA purification.

5. PRINT RNA synthesis

5.1 Before you begin

Timing: 1–2 days: Unless using a phage RNAP alternative, ensure that the designed plasmid templates for IVT of PRINT mRNA and template RNAs contain a T7 RNAP promoter immediately followed by one or more guanosines (5′-

TAATACGACTCACTATA(G)$_n$-3′). If linearizing templates for run-off transcription, ensure that the restriction enzyme recognition sequence is present at the intended run-off point and not also in the transcribed region. Positive control plasmids for transcription of PRINT template RNAs encoding an expression cassette for green fluorescent protein (GFP) or mCherry are available from Addgene (https://www.addgene.org/browse/article/28238145/). All materials used should be RNase-free and care should be taken to avoid RNase contamination.

5.2 Equipment

- Thermocycler
- Microcentrifuge (refrigerated)
- Nanodrop spectrophotometer
- Agarose gel electrophoresis set-up
- Urea-PAGE set-up
- P2 micropipette for pipetting volumes less than 2 μL

5.3 Recommended reagents and kits

- High-fidelity Type IIS restriction enzyme (e.g. in our design *Bbs*I-HF, NEB) and restriction digest buffer (supplied by manufacturer)
- QIAquick PCR Purification kit (Qiagen) or similar PCR clean-up kit
- HiScribe T7 High-Yield RNA Synthesis kit (NEB)
- 10X AG-cap Transcription Buffer (400 mM Tris-HCl (pH 8), 100 mM DTT*, 20 mM spermidine, 0.02% Triton X-100, 165 mM magnesium acetate, and nuclease-free H_2O) - store in aliquots at −80 °C

  *Spike in DTT fresh before use
- Ribonucleotide triphosphates (NTPs) (included in HiScribe kit) and recommended modified NTPs (m1ΨTP, ΨTP) (TriLink Biotechnologies)
- CleanCap AG (TriLink Biotechnologies)
- Murine RNase inhibitor (NEB)
- Yeast Inorganic Pyrophosphatase (NEB)
- T7 RNA Polymerase (NEB) (note that this component is distinct from the T7 RNA Polymerase mix included in the HiScribe kit listed above)
- RNase-free DNase I (ThermoFisher) and $MgCl_2$ buffer (supplied by manufacturer)
- Amersham Microspin G-25 columns (Cytiva)
- Phenol-chloroform-isoamyl alcohol (25:24:1 v/v, pH 6.7)
- Lithium chloride precipitation solution (7.5 M, ThermoFisher)

- Ethanol
- Sodium citrate (pH 6.5)
- 2X formamide-EDTA loading buffer (95% formamide, 5 mM EDTA, 0.025% bromophenol blue, 0.025% xylene cyanol)
- SYBR Gold (ThermoFisher)

5.4 Procedure

1. Linearize plasmids for R2p mRNA and template RNA IVT by digestion with a high-fidelity Type IIS restriction enzyme such as *Bbs*I-HF (NEB) (recommended) or amplify a linear template by PCR. For restriction digest, 5 μg plasmid can be digested with 5 μL enzyme in a 50-μL reaction at 37 °C for at least 4 h. Complete digestion should be verified by agarose gel electrophoresis alongside an undigested control. Complete digestion is important because any plasmid that escapes linearization will be a preferred IVT template, and continuous transcription will generate a high yield of unwanted product.
2. Purify linearized plasmid, for example by using the column-based QIAquick PCR Purification kit (Qiagen) with a separate column for each 5 μg of digested plasmid DNA, with elution in 25–30 μL nuclease-free H_2O for each. The expected yield of DNA is approximately 100–200 ng/μL. If any DNA form other than the desired linearized IVT template is present, purification of the desired IVT template by agarose gel electrophoresis and gel slice extraction can be used at this step with some loss of yield.

   Stopping point: Linearized plasmids can be stored at −20 °C for up to several months.
3. Set up IVT reactions on ice. Suggested reaction volumes can be scaled up or down as needed. Modified NTPs can be substituted for unmodified NTPs at equivalent concentration.

   a. For R2p mRNA preparation, use the CleanCap AG protocol from TriLink Biotechnologies (https://www.trilinkbiotech.com/media/folio3/productattachments/product_insert/n7113_insert_v3.pdf). We recommend setting up at least two 20 μL reactions, which can later be pooled. Use approximately 1 μg linearized DNA template per 20-μL reaction.

Synthesis of PRINT mRNA

Component	Final concentration	Volume per 20-μL reaction
ATP (100 mM)	5 mM	1 μL
CTP (100 mM)	5 mM	1 μL
GTP (100 mM)	5 mM	1 μL

Synthesis of PRINT mRNA

Component	Final concentration	Volume per 20-µL reaction
UTP (100 mM)	5 mM	1 µL
Recommended: Substitute with m1ΨTP (100 mM) or ΨTP (100 mM)		
CleanCap AG (100 mM)	4 mM	0.8 µL
10X AG-cap Transcription Buffer	1X	2 µL
Linearized DNA template	50 ng/µL	1 µg (volume will vary)
Murine RNase Inhibitor (40U/µL)	1 unit/µL	0.5 µL
Yeast Inorganic Pyrophosphatase (0.1 U/µL)	0.002 U/µL	0.4 µL
T7 RNA Polymerase (50 U/µL)	8 U/µL	3.2 µL
Nuclease-free H_2O	-	Up to 20 µL

b. For template RNA preparation, instead use the HiScribe T7 High-Yield RNA Synthesis Kit (NEB). Use approximately 1 µg linearized DNA template per 20-µL reaction.

Synthesis of template RNA

Component	Final concentration	Volume per 20-µL reaction
ATP (100 mM)	10 mM	2 µL
CTP (100 mM)	10 mM	2 µL
GTP (100 mM)	10 mM	2 µL
UTP (100 mM)	10 mM	2 µL
Strongly recommended: Substitute with ΨTP (100 mM)		
10X HiScribe Reaction Buffer	1X	2 µL
DTT (0.1 M)	5 mM	1 µL
Linearized DNA template	50 ng/µL	1 µg (volume will vary)
T7 RNA Polymerase Mix	-	2 µL
Nuclease-free H2O	-	Up to 20 µL

4. The IVT reactions should be mixed thoroughly by pipetting or gentle vortexing and then incubated at 37 °C for 2 h. Shorter RNAs (<500 nt) may require a longer reaction time to maximize yield by increasing the number of transcription initiation events.

5. Enrich IVT RNA by DNA removal with RNase-free DNase I (ThermoFisher). To each 20-μL reaction, add 23 μL nuclease-free H_2O, 5 μL 10X $MgCl_2$ reaction buffer (ThermoFisher), and 2 μL DNase I (ThermoFisher) for a final reaction volume of 50 μL. Incubate at 37 °C for 30 min

6. Desalt each IVT RNA to remove transcription buffer and unincorporated NTPs. The unincorporated NTPs interfere with determination of purified RNA concentration. We use Amersham Microspin G-25 columns (Cytiva) for this step.

7. Purify IVT RNA.

    a. We use extraction with phenol-chloroform-isoamyl alcohol (25:24:1 v/v PCI, pH 6.7, ThermoFisher). Dilute each 50 μL reaction to 100 μL with nuclease-free H_2O. Add an equal volume (100 μL) PCI to each tube and vortex gently to mix. Centrifuge at 18,000 × g for 5 min at 4 °C. Transfer the aqueous layer to a fresh tube.

    b. Though untested, column-based RNA purification should work as a replacement. Skip step 8 if using this purification method.

8. Precipitate IVT RNAs with LiCl. Add ½ volume (50 μL) 7.5 M LiCl to each tube and vortex gently to mix. RNA can be precipitated by tube immersion in liquid N_2 or at −80 °C overnight. LiCl precipitation enriches for large RNAs and thus removes unwanted short T7 RNAP abortive initiation products and transcript leader sequences released by Rz self-cleavage, which could hinder PRINT efficiency by their 5′ triphosphate.

    Stopping point: After the addition of LiCl, reactions can be stored at −80 °C for several days.

9. Place the LiCl precipitation reactions at room temperature until thawed. Centrifuge at 18,000 × g for 30 min at 4 °C to pellet RNA. Wash the pellet twice with ice-cold 75% EtOH. Let the RNA pellet air dry for approximately 20 min or until translucent.

10. Resuspend RNA in 1 mM sodium citrate (pH 6.5) on ice for at least 30 min, vortexing occasionally on low speed. R2p mRNAs will have a lower yield than template RNA; two pooled 20-μL reactions can be resuspended in 30 μL for an expected yield of approximately 1–2 μg/μL. Template RNAs from a single 20-μL reaction can be resuspended in 50 μL for an expected yield of approximately 1–2 μg/μL. Note that use of uridine analogs may result in a slight decrease in yield.

11. Record yield and store. We assess RNA concentration by Nanodrop spectrophotometry. If you take a multi-wavelength measurement, note that modification with uridine analogs may alter typical 260/280 (expected: >2.0) and 260/230 (expected: ~2.0–2.3) values for RNA. Aliquot RNA and store at −80 °C to avoid repeated freeze-thaws, which compromise RNA integrity. Save an appropriate amount (50–100 ng) for quality control.

12. Check RNA quality.

a. Run approximately 25–50 ng of purified RNAs on a denaturing 5% Urea-PAGE (19:1 Acrylamide/Bis, 7 M Urea) gel. Note that use of gels purchased from a commercial vendor may require optimization from these conditions. Dilute each RNA stock appropriately (typically 1:100) in nuclease-free H_2O and combine with an equal volume of 2X formamide-EDTA loading buffer. Heat RNA at 65 °C for 3 min, then immediately move to ice before loading. Following electrophoresis, gels can be stained using SYBR Gold (ThermoFisher) for 15 min at room temperature. Note that long RNAs may retain structure that retards mobility, and this effect is exacerbated by incorporation of m1ΨTP or ΨTP. Smearing is indicative of degraded RNA.

b. Other size profiling methods capable of sizing long RNAs could serve as a replacement for this step.

6. Co-transfection of PRINT RNAs and cell harvest

6.1 Before you begin

Timing: 3 days: All steps here assume reverse transfection of adherent cells in 6-well plate format. This protocol has been specifically optimized for hTERT RPE-1 cells, which have high PRINT efficiency (Zhang et al., 2024). We maintain hTERT RPE-1 cells by serial passaging up to 20 passages in DMEM/F12 (Gibco) supplemented with 10% fetal bovine serum (FBS) and 100 μg/mL Primocin (Invivogen) (henceforth referred to as complete medium). Cells are incubated at 37°C under 5% CO_2. One well of transfected hTERT RPE-1 cells is sufficient for analysis by the combination of flow cytometry, PCR genotyping, and droplet digital PCR (ddPCR). Transfections may be scaled up or down for other applications. We recommend carrying out technical triplicates of all transfections. The addition of mCherry or GFP mRNA as a spike-in to the PRINT RNA mixture can be used for normalization of flow cytometry data.

6.2 Equipment

- Biosafety cabinet
- Cell culture incubator
- Table-top centrifuge (50-mL conical tube capacity)
- P2 micropipette for pipetting volumes less than 2 μL
- Hemocytometer or cell counter
- Flow cytometer (if applicable)

6.3 Reagents

- Dulbecco’s phosphate-buffered saline (DPBS, Gibco)
- Trypsin-EDTA (0.05%, Gibco)

- DMEM/F12 (Gibco) supplemented with 10% FBS and 100 μg/mL Primocin (Invivogen)
- OptiMEM (Gibco)
- Lipofectamine MessengerMax Transfection Reagent (ThermoFisher)
- DPBS (Gibco) supplemented with 2% FBS and 0.5 mM EDTA (if performing flow cytometry)
- (Optional) mCherry or GFP mRNA spike-in transfection control (commercial option: TriLink catalog number L-7601 or L-7203)

6.4 Procedure

1. Prepare cells one day prior to transfection. Split cells such that they will be in log-phase of growth (30–50% confluent) at the time of transfection. Cells in stationary phase will have significantly reduced PRINT efficiency.
2. Prior to transfection, thaw IVT RNAs on ice. Calculate how much of each RNA to transfect. We recommend transfecting 1 μg total RNA per well of a 6-well plate at an optimal R2p mRNA to template RNA molar ratio of 1:3. The amount of total RNA transfected can be increased up to 2.5 μg, but higher RNA dosage will increase cell death. Each R2p mRNA and template RNA used should also be transfected alone as a negative control.
3. Pre-warm complete medium in wells to be transfected by adding 1 mL to each well of a 6-well plate and incubating at 37 °C while assembling transfections.
4. Prepare cells for transfection immediately prior to use. Aspirate media, wash once with DPBS (Gibco), and release cells from plate with trypsin by incubation at 37 °C for 5 min. Add complete medium to cells to inactivate trypsin and pipette up and down to break up cell clumps. Count cell density using a hemocytometer or cell counter. Pellet cells and resuspend in complete medium to a volume such that 1 mL of cell suspension yields the number of cells desired per well. For hTERT RPE-1 cells, we recommend using 750,000–1,000,000 cells per well of a 6-well plate; this may be scaled up or down as needed and may need to be adjusted based on cell type.
5. Prepare two microcentrifuge tubes per transfection. In one tube, add 125 μL of OptiMEM (Gibco) and 2 μL (i.e. what will be a 1:2 RNA to transfection reagent mass ratio) of Lipofectamine MessengerMax (ThermoFisher). Pipette up and down to mix. Let incubate for 10 min at room temperature. Using a higher mass ratio of MessengerMax to RNA will increase transfection efficiency but also result in higher cell death.
6. While MessengerMax is incubating, add the appropriate volume of IVT RNAs to the second tube on ice. If using a spike-in mRNA as a control for transfection efficiency (recommended if read-out is assessed by flow cytometry), mCherry or GFP mRNA (10 ng/1 million cells) can be added to the IVT RNA mixture at this point. After addition of all IVT RNA components, briefly spin down the tube.

7. Following the 10 min MessengerMax incubation, add 125 μL of OptiMEM to the IVT RNAs and pipette up and down to mix. Add RNA mixture to MessengerMax mixture and pipette up and down to mix. Let incubate for 5 min at room temperature.

8. Resuspend cells in their medium by gentle inversion (cells will settle during preparation of RNA and MessengerMax) and add 1 mL cell suspension to pre-warmed media per well of a 6-well plate. Add MessengerMax-RNA mixture dropwise to cells and rock plates gently to mix. Incubate at 37 °C.

9. Optional: Aspirate media containing MessengerMax 4–6 h post-transfection and replace with 2 mL fresh complete medium per well. Incubate at 37 °C until harvest.

10. To harvest adherent cells for flow cytometry analysis, aspirate media, wash once with DPBS, and trypsinize as described above. Resuspend cells in 1 mL DPBS supplemented with 2% fetal bovine serum (FBS) and 0.5 mM EDTA. Pipette up and down to break up cell clumps and transfer to a flow cytometry tube. Keep cells on ice until flow cytometry analysis. Flow cytometer use will vary by instrument.

11. Following or instead of flow cytometry, cell pellets may be kept for downstream analysis by immunoblotting, PCR genotyping, ddPCR, high-throughput sequencing, or other analysis. Centrifuge cells at 7000 × g for 3 min, wash once with DPBS, snap-freeze pellets in liquid N_2, and store at −80 °C.

7. Analysis of PRINTed cells

Several methods can be used to analyze transgene expression, insertion efficiency, and site-specificity of insertion following a PRINT experiment (Zhang et al., 2024). Common methods are indicated in Fig. 6A. In the case in which a fluorescent protein-encoding transgene is PRINTed, transgene expression can be analyzed by flow cytometry. Flow cytometry provides a direct readout of the percentage of cells expressing the transgene (see Fig. 6B for an example). Additionally, median fluorescence intensity of cells expressing the fluorescent transgene provides an approximation of relative transgene copy number (Fig. 6B). In PRINT experiments analyzed by flow cytometry, it is critical to include a template RNA-only control for each template RNA used, or more stringently, template RNA and RTD R2p mRNA. The template alone may yield a low percentage of cells that gate as positive for low-intensity fluorescence, and these must be subtracted as background to accurately assess PRINT transgene expression. We note that the relative size of this population varies with cell type but is typically less than 1% of analyzed cells. Transfection of R2p mRNA alone provides an indication of possible toxicity.

In PRINTed hTERT RPE-1 cells, expression of a GFP transgene as assessed by flow cytometry is readily detectable at 6 h post-transfection and becomes maximal within 24 h (Zhang et al., 2024). Therefore, we recommend performing flow cytometry analysis 20–24 h post-transfection for general assessment of PRINT efficiency. In cases in which a

non-fluorescent protein-encoding transgene is PRINTed, alternative methods can be used to detect transgene expression, such as immunoblotting.

In conjunction with flow cytometry, additional approaches may be employed to assess the efficiency and accuracy of PRINT experiments. PCR genotyping of genomic DNA (gDNA) extracted from PRINTed cells is a simple way to verify that a transgene has been successfully inserted into the 28S rDNA locus (see Fig. 6C for an example). Primers specific to 28S rDNA and the inserted transgene (see Table 2) are used to amplify rDNA-transgene 5′ and 3′ junctions by PCR (Fig. 6C, top), and products are detected on an agarose gel (Fig. 6C, bottom). Detection of products at the expected size from PRINTed cells but not cells transfected with R2p mRNA only or template RNA only controls is indicative of successful PRINT. The relative amount of product amplified also provides a semi-quantitative measurement of inserted transgene copy number.

To more precisely quantify transgene copy number, ddPCR analysis of gDNA from PRINTed cells may be useful. In this method, transgene copy number is calculated relative to a reference gene in end-point PCR reactions that have been partitioned into thousands of individual small droplets. The high sensitivity and quantitative read-out of ddPCR allows precise determination of copy number across transgene segments, which generally decreases from 3′ to 5′ due to transgene 5′ truncations. The percentage of full-length transgenes in a population of PRINTed cells can be determined by comparing copy number at the 3′ and 5′ ends of the transgene. While transgene 3′ copy number can be determined using a ddPCR probe which spans the precise rDNA-transgene 3′ junction, ddPCR probes for the transgene 5′ end should not span the 5′ junction because 5′ junction sequences are more heterogeneous. Instead, we recommend choosing a target amplicon within the template RNA 5′ module or transgene RNAP II promoter close to the transgene 5′ end. In addition, the site-specificity of transgene insertions can be determined by whole-genome sequencing (WGS) analysis. By this method, we determined that even the initial PRINT technology has high target-site precision with fewer than 1% off-target insertions (Zhang et al., 2024).

8. Future optimizations and applications

The protocol detailed here offers a sturdy foundation for the continued optimization and application of PRINT. As with any new technology, there are ample opportunities to better understand biological requirements for transgene insertion and to improve the method efficiency, specificity, and percentage of productive insertions. Much remains unknown about mechanisms that support endogenous non-LTR retrotransposon mobility, so continued investigation into this biology will also inform PRINT. Combining PRINT with different delivery options will expand the repertoire of cell types that can be effectively PRINTed as well as open transgene delivery opportunities to in vivo models. A better understanding of host factor influence on PRINT will enable fine-tuned control of host response and permit reliably full-length insertions. More stringent purification of R2p mRNA and template RNA will be useful in pushing the limits of inserted transgene length. Ultimately, PRINT is a promising avenue for the incorporation of therapeutically relevant genes into rDNA safe-harbor loci.

References

Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, … Macfarlan TS (2018). Ten things you should know about transposable elements. Genome Biology, 19, 199. [PubMed: 30454069]

Burke WD, Malik HS, Lathe WC, & Eickbush TH (1998). Are retrotransposons long-term hitchhikers? Nature, 392, 141–142. [PubMed: 9515960]

Chadalavada DM, Knudsen SM, Nakano S, & Bevilacqua PC (2000). A role for upstream RNA structure in facilitating the catalytic fold of the genomic hepatitis delta virus ribozyme. Journal of Molecular Biology, 301, 349–367. [PubMed: 10926514]

Chheda U, Pradeepan S, Esposito E, Strezsak S, Fernandez-Delgado O, & Kranz J (2024). Factors affecting stability of RNA - Temperature, length, concentration, pH, and buffering species. Journal of Pharmaceutical Sciences, 113(2), 377–385. [PubMed: 38042343]

Eickbush DG, Burke WD, & Eickbush TH (2013). Evolution of the R2 retrotransposon ribozyme and its self-cleavage site ed. M.A. Batzer. PLoS One, 8, e66441. [PubMed: 24066021]

Eickbush DG, & Eickbush TH (2003). Transcription of endogenous and exogenous R2 elements in the rRNA gene locus of *Drosophila melanogaster*. Molecular and Cellular Biology, 23(11), 3825–3836. [PubMed: 12748285]

Eickbush TH, & Eickbush DG (2015). Integration, regulation, and long-term stability of R2 retrotransposons. Microbiology Spectrum, 3(2), MDNA3-0011-2014.

Ferré-D′Amaré AR, Zhou K, & Doudna JA (1998). Crystal structure of a hepatitis delta virus ribozyme. Nature, 395, 567–574. [PubMed: 9783582]

Hacein-Bey-Abina S, Garrigue A, Wang GP, Soulier J, Lim A, Morillon E, … Beldjord K (2008). Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. The Journal of Clinical Investigation, 118(9), 3132–3142. [PubMed: 18688285]

Jakubczak JL, Burke WD, & Eickbush TH (1991). Retrotransposable elements R1 and R2 interruupt the rRNA genes of most insects. PNAS, 88, 3295–3299. [PubMed: 1849649]

Karikó K, Muramatsu H, Welsh FA, Ludwig J, Kato H, Akira S, & Weissman D (2008). Incorporation of pseudouridine into mRNA yields superior nonimmunogenic vector with increased translational capacity and biological stability. Molecular Therapy, 16, 1833–1840. [PubMed: 18797453]

Kojima KK, & Fujiwara H (2005). Inheritance of the 28S rDNA-Specific Retrotransposon R2. Molecular Biology and Evolution, 22(11), 2157–2165. [PubMed: 16014872]

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(9), e0163496. [PubMed: 27662593]

Lee RJ, Horton CA, Van Treeck B, McIntyre JJR, & Collins K (2024). Conserved and divergent DNA recognition specificities and functions of R2 retrotransposon N-terminal domains. Cell Reports, 43, 114239. [PubMed: 38753487]

Lévesque D, Choufani S, & Perreault J-P (2002). Delta ribozyme benefits from a good stability in vitro that becomes outstanding in vivo. RNA, 8, 464–477. [PubMed: 11991641]

Li T, Yang Y, Qi H, Cui W, Zhang L, Fu X, … Yu T (2023). CRISPR/Cas9 therapeutics: Progress and prospects. Signal Transduction and Targeted Therapy, 8, 36. 10.1038/s41392-023-01309-7. [PubMed: 36646687]

Luan DD, Korman MH, Jakuubczak JL, & Eickbush TH (1993). Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: A mechanism for nonn-LTR retrotransposition. Cell, 72(4), 595–605. [PubMed: 7679954]

Luchetti A, & Mantovani B (2013). Non-LTR R2 element evolutionary patterns: Phylogenetic incongruences, rapid radiation and the maintenance of multiple lineages. PLoS One, 8(2), e57076. [PubMed: 23451148]

Palm SM, Horton CA, Zhang X, & Collins K (2024). Structure and sequence at an RNA template 5′ end influence insertion of transgenes by an R2 retrotransposon protein. RNA, 30(9), 1227–1245. [PubMed: 38960642]

Pichlmair A, Schulz O, Tan CP, Näslund TI, Liljeström P, Weber F, … Sousa C (2006). RIG-I-mediated antiviral responses to single-stranded RNA bearing 5′-phosphates. Science, 314, 997–1001. [PubMed: 17038589]

Sandoval-Villegas N, Nurieva W, Amberger M, & Ivics Z (2021). Contemporary transposon tools: A review and guide through mechanisms and applications of sleeping beauty, piggyBac and Tol2 for genome engineering. International Journal of Molecular Sciences, 22(10), 5084. [PubMed: 34064900]

To KKW, & Cho WCS (2021). An overview of rational design of mRNA-based therapeutics and vaccines. Expert Opinion on Drug Discovery, 16, 1307–1317. [PubMed: 34058918]

Trepotec Z, Geiger J, Plank C, Aneja MK, & Rudolph C (2019). Segmented poly (A) tails significantly reduce recombination of plasmid DNA without affecting mRNA translation efficiency or half-life. RNA, 25, 507–518. [PubMed: 30647100]

Wells JN, & Feschotte C (2020). A field guide to eukaryotic transposable elements. Annual Review of Genetics, 54, 539–561.

World Health Organization. (2020). Messenger RNA encoding the full-length SARS-CoV-2 spike glycoprotein. Nature Biotechnology, https://mednet-communities.net/inn/db/media/docs/11889.doc.

Yang J, & Eickbush TH (1998). RNA-induced changes in the activity of the endonuclease encoded by the R2 retrotransposable element. Molecular and Cellular Biology, 18(16), 3455–3465. [PubMed: 9584185]

Yang H, Ren S, Yu S, Pan H, Li T, Ge S, … Xia N (2020). Methods favoring homology-directed repair choice in response to CRISPR/Cas9 induced-double strand breaks. International Journal of Molecular Sciences, 21(18), 6461. 10.3390/ijms21186461. [PubMed: 32899704]

Zhang X, Van Treeck B, Horton C, McIntyre J, Palm S, Shumate J, & Collins K (2024). Harnessing eukaryotic retroelement proteins for transgene insertion into human safe-harbor loci. Nature Biotechnology. 10.1038/s41587-024-02137-y.

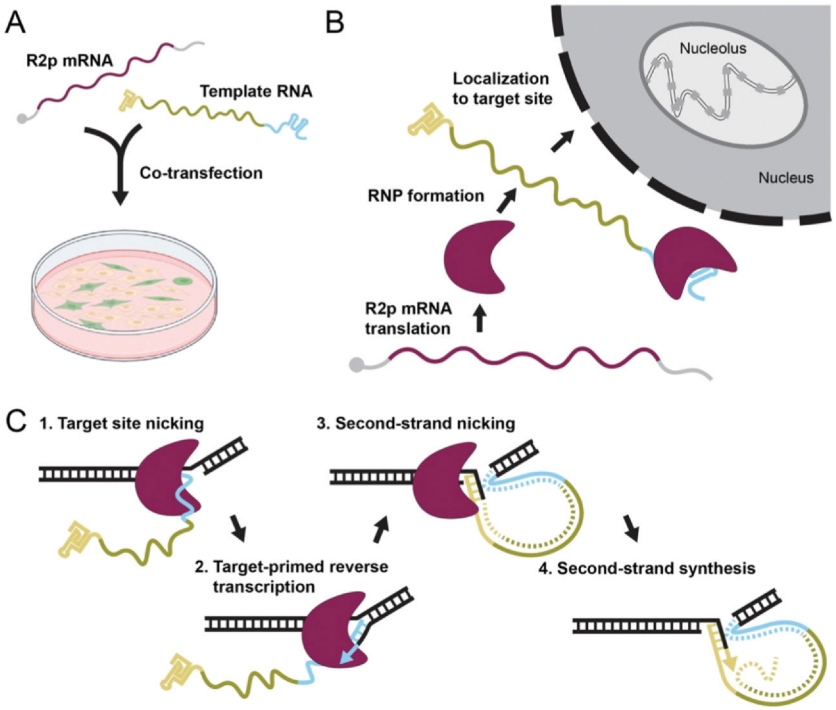

Fig. 1. Overview of precise RNA-mediated insertion of transgenes (PRINT).

(A) PRINT relies on the transfection of two RNAs: an mRNA encoding R2p and a transgene-encoding template RNA. (B) Upon successful delivery to the cell, the mRNA is translated into R2p. R2p recognizes the R2 3′ UTR sequence at the 3′ end of the transgene template RNA. The RNP then finds the highly conserved target site within the repetitive rDNA arrays compartmentalized in the nucleolus. (C) Insertion mechanism utilized in PRINT. One strand of the target site is nicked. Then, this nicked strand is used to prime cDNA synthesis of the transgene directly into the genome through target-primed reverse transcription (TPRT). Nicking and synthesis of the second strand complete insertion.

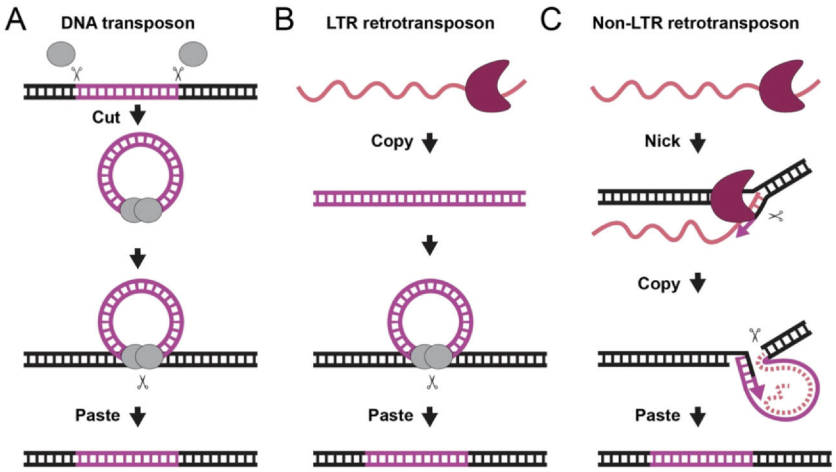

Fig. 2. Transposable elements differ in their mechanisms of insertion.

(A) DNA transposons insert via a “cut and paste” mechanism. The DNA is excised from one site and planted at a new insertion site. (B) LTR retrotransposons insert via a “copy and paste” mechanism. The retroelement RNA is reverse-transcribed to create extrachromosomal double-stranded DNA that is then integrated into the genome. (C) Non-LTR retrotransposons insert via a “nick-copy-paste” mechanism. This mechanism avoids the creation of extrachromosomal DNA as the reverse transcription is primed from nicked genomic DNA.

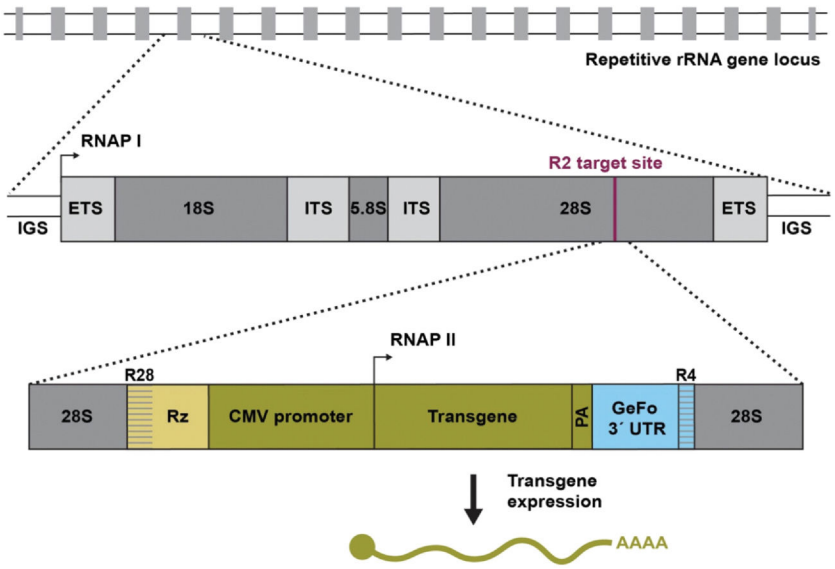

Fig. 3. A safe-harbor locus within rDNA.

The R2 retrotransposon target site is a highly conserved sequence within the 28S rDNA sequence. The presence of hundreds of copies of the target site allows the tolerance of insertion of a handful of transgenes. Successfully inserted transgenes harbor an RNAP II promoter sequence for canonical mRNA transcription, processing, and translation.

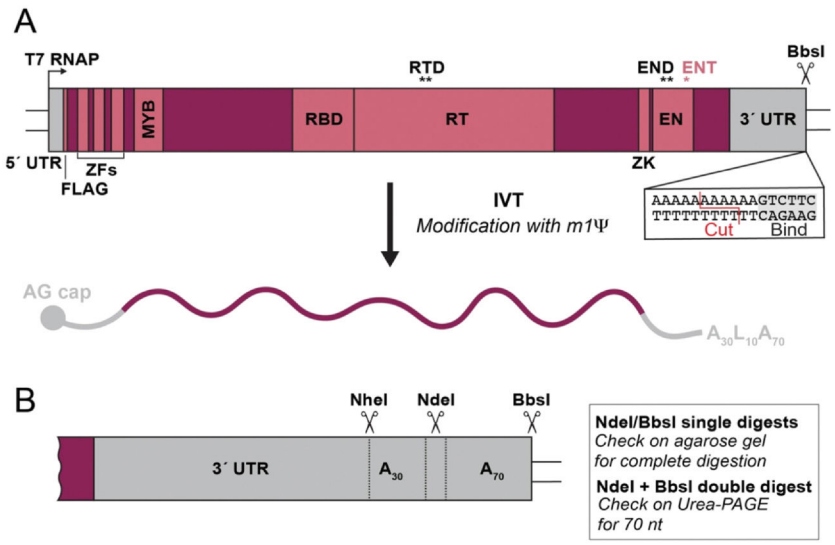

Fig. 4. R2 mRNA design.

(A) R2p has all activities required for site-specific TPRT encoded as one protein by a single ORF (magenta). *ZFs*, zinc fingers; *MYB*, Myb DNA-binding domain; *RBD*, RNA-binding domain; *RT*, reverse transcriptase domain; *ZK*, zinc knuckle; *EN*, endonuclease domain. An optional FLAG tag resides at the N-terminus. The BioNTech 5′ and 3′ UTRs flank the ORF. IVT of a *Bbs*I-digested R2p plasmid results in mRNA with an AG cap, optimally with a uridine analog in replacement of uridine. Asterisks denote the approximate locations of mutations for RTD, END and ENT versions of the proteins described in the main text. Inset: *Bbs*I is a Type IIS restriction enzyme that cuts outside of its binding site. (B) Zoom-in from the 3′ end of A. $A_{30}L_{10}A_{70}$ is unstable during plasmid replication. Restriction enzyme digestion sites that can be used for quality checks of full sequence retention within the plasmid are indicated.

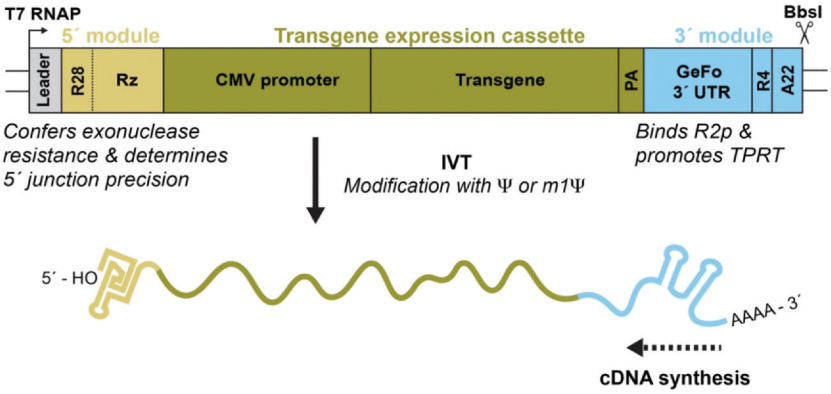

Fig. 5. Depiction of optimized template RNA for use in PRINT.

Top, a DNA template for synthesis of template RNA contains the template RNA components required for efficient PRINT, a preceding T7 RNAP promoter, and a recognition site for a high-fidelity Type IIS restriction enzyme. The DNA template is linearized by the Type IIS restriction enzyme then used for IVT of the PRINT template RNA (bottom).

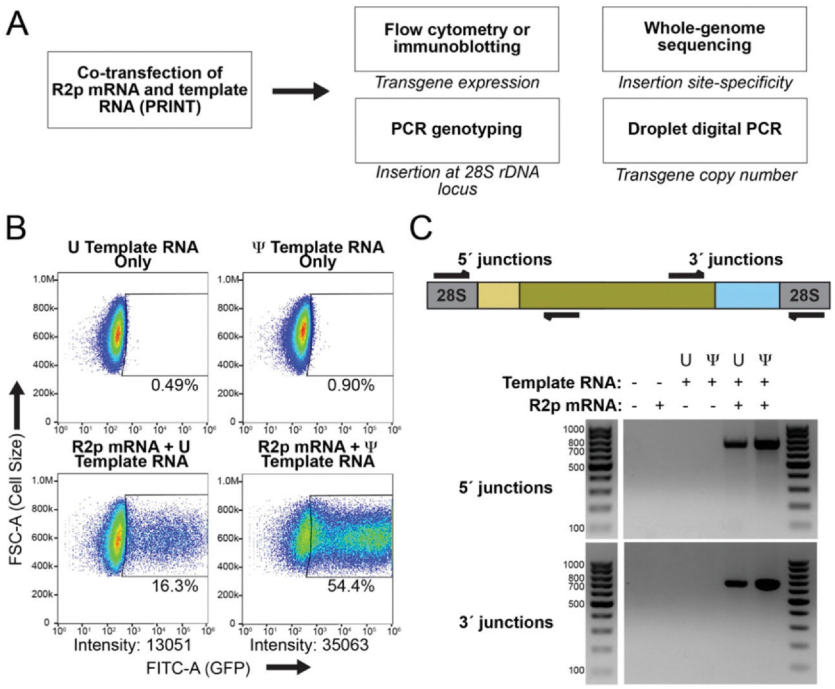

Fig. 6. Methods for analysis of PRINTed cells.

(A) Common methods used for analysis of transgene insertion and expression following PRINT. (B) An example of flow cytometry analysis of hTERT RPE-1 cells following transfection of a GFP-encoding PRINT template RNA alone (top) or co-transfection with R2p mRNA (bottom). Template RNAs were synthesized with uridine or Ψ. The percentage of cells expressing GFP is determined by the percentage of cells gated as GFP-positive for the co-transfection condition minus those gated as GFP-positive for the template RNA alone condition. Median intensity of GFP+ cells (indicated below the x-axis for the two PRINT conditions) can be used as a proxy for insertion copy number. (C) An example of PCR genotyping analysis using gDNA extracted from hTERT RPE-1 cells following PRINT. Top, a schematic showing PCR amplification of 28S rDNA-transgene 5′ and 3′ junctions. Bottom, a representative agarose gel with 5′ and 3′ junction PCR products.

Table 1

28S rDNA target site sequence for PRINT insertion.

	28S Target Site Sequence
Avian/Human	5′-AGCGCGGGTAAACGGCGGGAGTAACTAT GACTCTCTTAAG**GT**AGCCAAATGCCTCGTCAT-3′

The insertion occurs between the two nt in bold.

Table 2

Oligonucleotides for 5′ and 3′ rDNA-transgene junction PCR amplification of gDNA from PRINTed cells.

Primer	Region amplified	Sequence
28S rDNA FW	5′ junctions	5′-CCAGGGGAATCCGAC TGTTTAATTAAAACAAAGC-3′
TriCasA Rz REV	5′ junctions	5′-CCCTGTGCCCAGCGGGAC-3′
CMV REV	5′ junctions	5′-GGGCGTACTTGGCATATGA TACACTTGATG-3′
GFP FW	3′ junctions	5′-CACTCTCGGCATGGACGAG CTGTAC-3′
GeFo3 FW	3′ junctions	5′-GGACGGGCCACC TTTACTTAACCC-3′
28S rDNA REV	3′ junctions	5′-CAGACTAGAGTC AAGCTCAACAGGGTCTTC-3′

FW is forward (rDNA sense strand), and REV is reverse (rDNA antisense strand). The GFP FWD primer is in the GFP coding region.