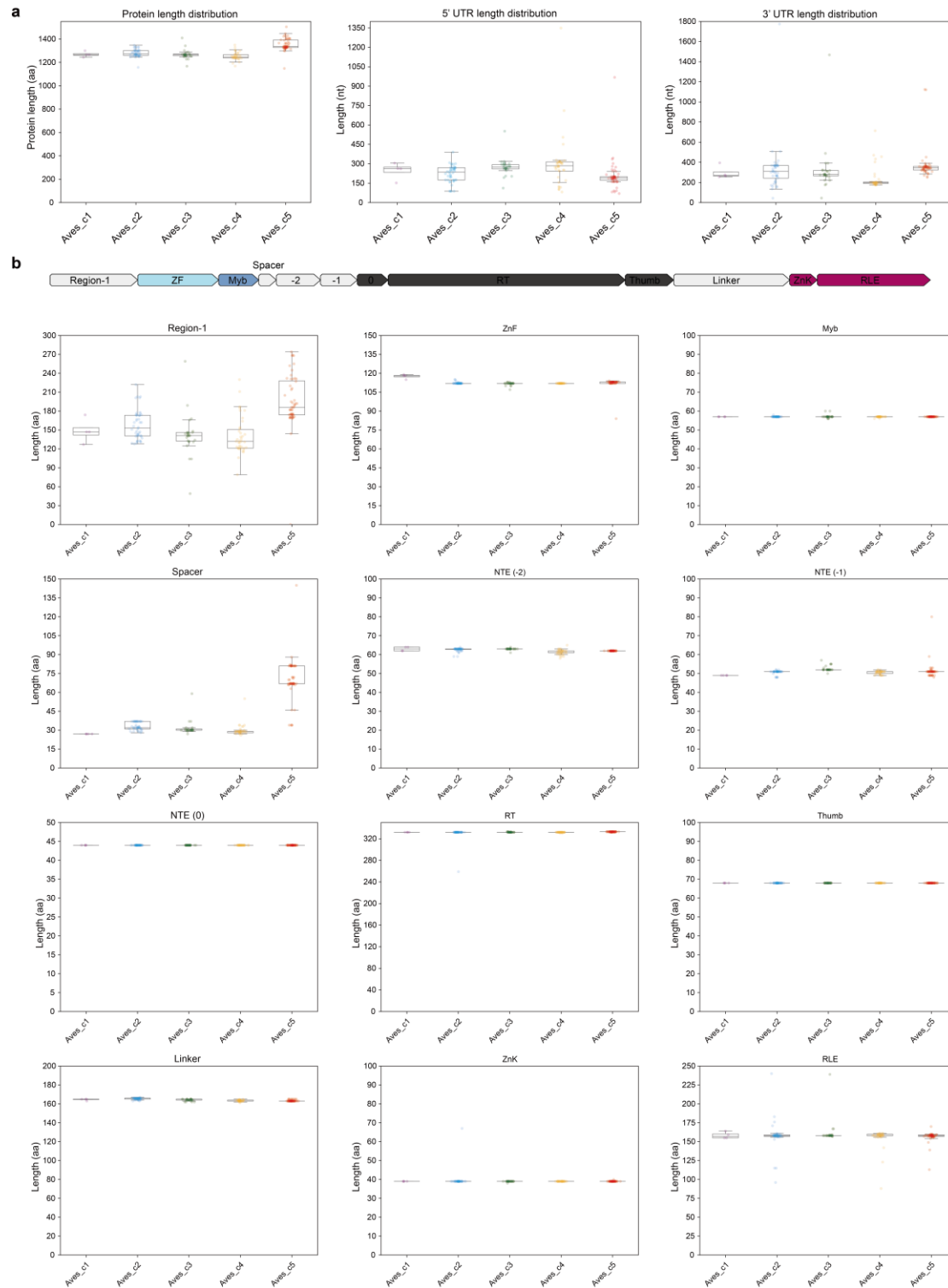


Discovery and engineering of avian R2 retrotransposons for all-RNA-mediated targeted DNA integration in human cells

In the format provided by the
authors and unedited

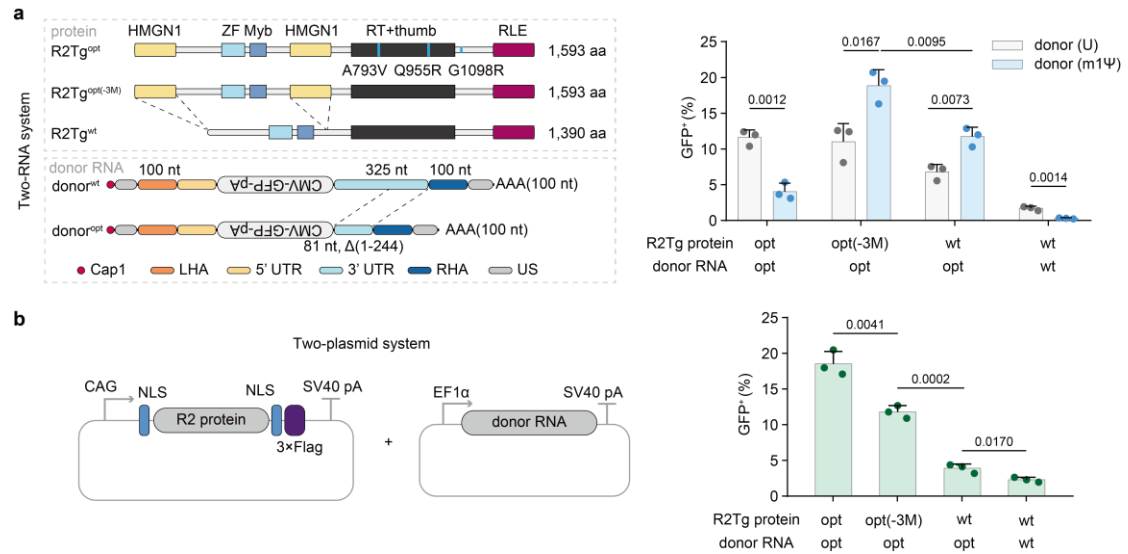
List of Supplementary Information:
Supplementary Figure 1 to 9



Supplementary Fig. 1| Statistical analysis of R2 retrotransposons from Aves_c1-c5

a, Length distributions of proteins, 5' UTRs, and 3' UTRs in R2 elements from Aves_c1-c5 (n = 4, 44, 27, 38, and 52, respectively). Box centers indicate medians; box limits indicate 25th and 75th percentiles; whiskers extend to minimum and maximum values.

b, Statistical analysis of domain lengths in R2 proteins from Aves_c1-c5.



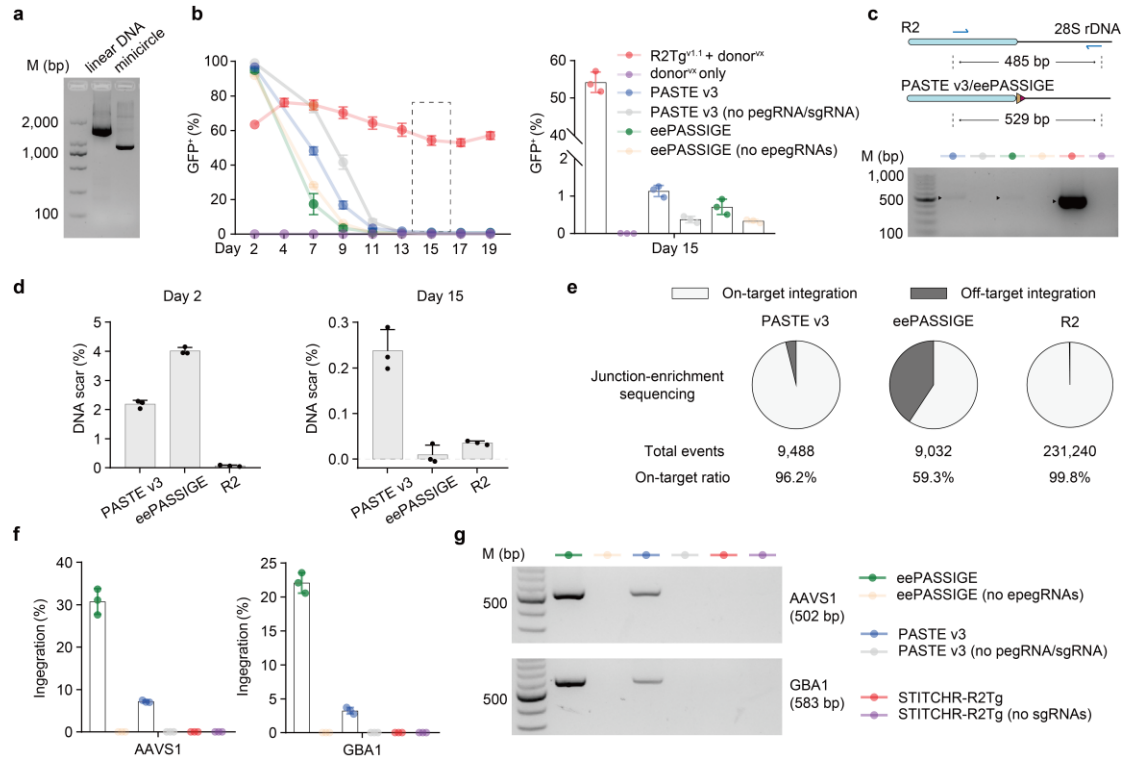
Supplementary Fig. 2 | Investigation of incompatibility between en-R2Tg and m1Ψ-modified RNA donors

a, Schematic of the two-RNA system (R2 protein + donor RNA). R2Tg^{opt(-3M)} refers to R2Tg^{opt} with three engineered point mutations removed¹. Integration efficiency of different R2Tg proteins with unmodified or m1Ψ-modified RNA donors was evaluated in HEK293T cells. LHA, left homologous arm; RHA, right homologous arm; US, unrelated vector sequence.

b, Integration efficiency of a two-plasmid system (expressing R2 protein and donor RNA) in HEK293T cells.

Unless otherwise stated, data are presented as mean ± standard deviation (SD) from three biological replicates.

Unpaired, two-tailed Student's t tests.



Supplementary Fig. 3 | Systematic comparison of the engineered R2 tool developed in this study with DNA donor-based gene integration tools (PASTE v3 and eePASSIGE) in terms of activity and safety

a, Agarose gel electrophoresis of minicircle DNA donors for PASTE v3 and eePASSIGE.

b, Long-term cell culture of HEK293T cells treated with PASTE v3, eePASSIGE or the engineered R2 tool. Minicircles harbored a CMV-GFP expression cassette, resulting in high initial GFP⁺ cell percentages for PASTE v3 and eePASSIGE groups; GFP⁺ cell percentages stabilized after Day 15, with experimental groups remaining significantly higher values than the control group.

c, Genotyping analysis of samples collected on Day 15 from Supplementary Fig. 3b. This experiment was repeated independently three times with similar results. Representative gel image from one experiment was shown.

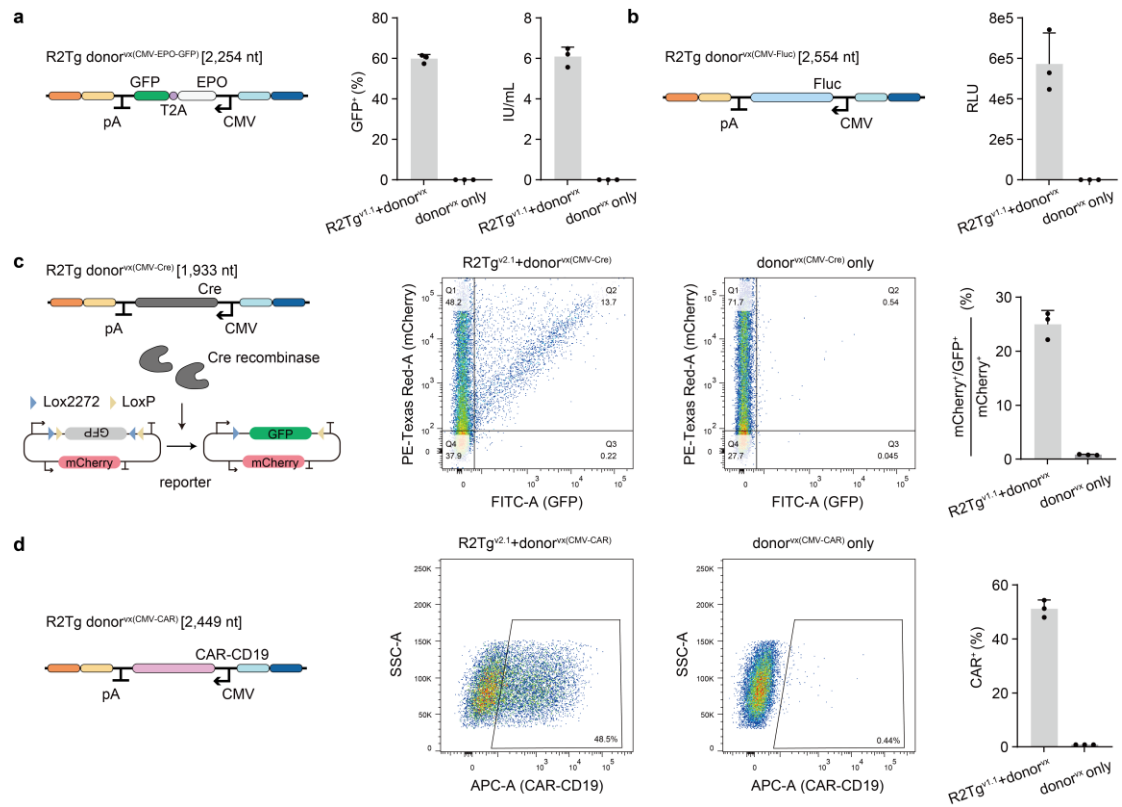
d, Changes in DNA scar levels on days 2 and 15 across three technologies. DNA scars represent gene editing byproducts unrelated to targeted transgene integration, including targeted indel frequency (defined as the proportion of rDNA loci containing indels relative to the total rDNA pool) and the proportion of AttP sites inserted via prime editing. DNA scar percentages were normalized by subtracting background values from the corresponding negative controls. These results suggest that while PASTE v3 and eePASSIGE may mediate efficient transgene integration at Day 2, the substantial rDNA damage they induced drives robust rDNA recombination (referring to Extended Data Fig. 8), leading to a marked decline in apparent integration efficiency by Day 15.

e, Unbiased assessment of gene integration specificity for the three technologies via junction-enrichment sequencing.

f, Assessment gene integration efficiency at non-rDNA sites for PASTE v3, eePASSIGE and STITCHR.

g, Genotyping analysis of samples collected from Supplementary Fig. 3f. This experiment was repeated independently three times with similar results. Representative gel image from one experiment was shown.

For all graphs, data are represented as mean $\pm$ SD from three biological replicates.



Supplementary Fig. 4 | Engineered R2Tg mediates efficient insertion of non-GFP genes in HEK293T cells

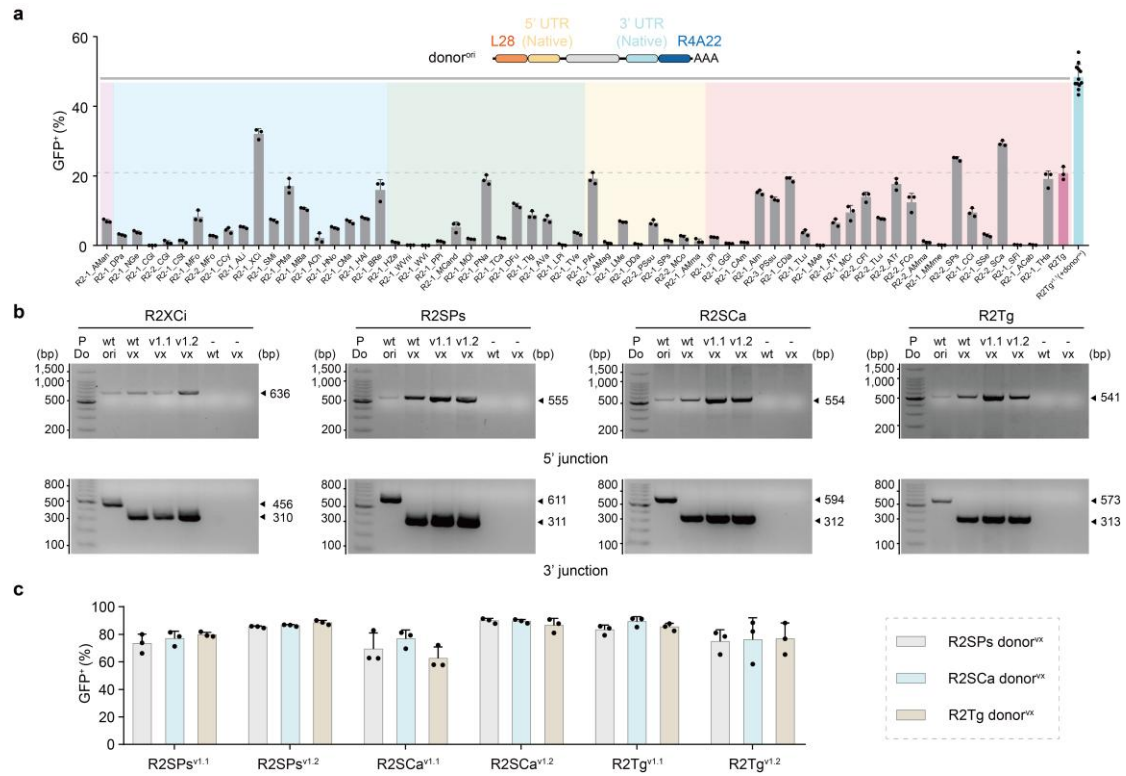
a, R2Tg^{v1.1}-mediated insertion of the EPO-T2A-GFP gene, with efficient EPO expression confirmed.

b, R2Tg-mediated insertion of the Firefly luciferase gene.

c, Evaluation of engineered R2Tg efficiency for inserting the Cre gene in HEK293T cells.

d, Evaluation of engineered R2Tg efficiency for inserting the CAR-CD19 gene in HEK293T cells.

For all graphs, data are represented as mean $\pm$ SD from three biological replicates.



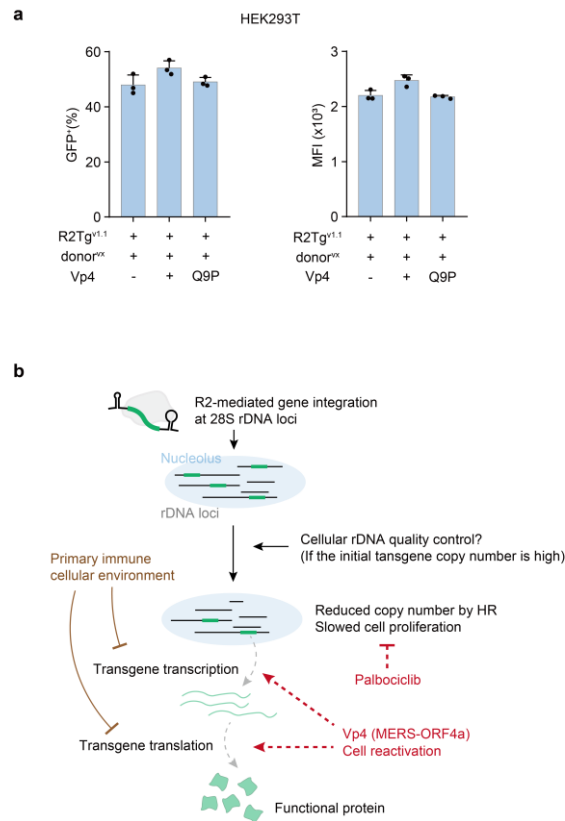
Supplementary Fig. 5 | Evaluating the gene integration efficiency of novel R2 orthologs in human HEK293T cells

a, Activity screening of new R2 proteins in HEK293T cells using donor^{ori} (donor^{vx} with native 3' UTR); R2Tg^{v1.1} + donor^{vx} served as a positive control. Lipo3000 was used as delivery reagent in this figure.

b, Agarose gel electrophoresis of amplified 5' and 3' junction fragments at 28S rDNA loci from cells treated with different R2 retrotransposons (corresponding to Fig. 2d). This experiment was repeated independently three times with similar results. Representative gel image from one experiment was shown.

c, Activity of R2 proteins paired with donors from different R2 species in HEK293T cells.

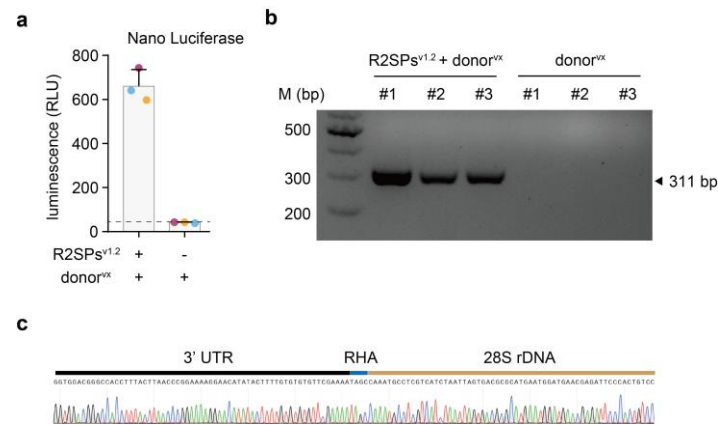
For all graphs, data are represented as mean $\pm$ SD from three biological replicates.



Supplementary Fig. 6 | Vp4 facilitates R2-mediated gene integration and a model for dynamic transgene regulation.

a, Effect of MERS-CoV ORF4a (Vp4) on R2-mediated transgene integration in HEK293T cells using R2Tg^{v1.1} and donor^{vx} carrying the CMV-EPO-T2A-GFP transgene. Data are presented as mean $\pm$ SD from three biological replicates.

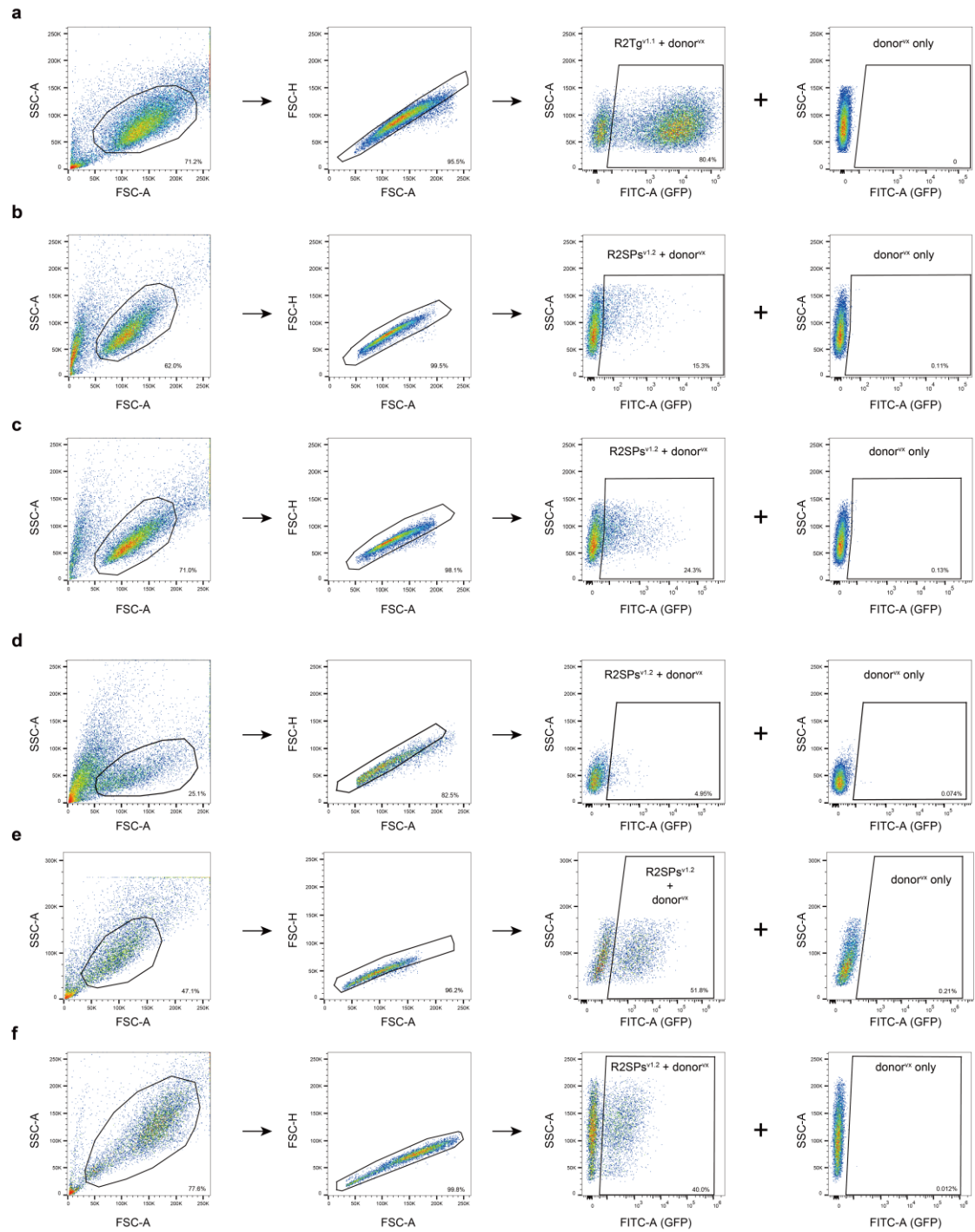
b, A model delineating the dynamic regulation of R2-retrotransposon-integrated transgenes at human rDNA loci across distinct cell types. This model proposes multi-layered regulation of transgenes integrated via R2 retrotransposons at human 28S rDNA loci across different human cell types. Engineered R2 retrotransposons first mediate site-specific integration of transgenes into rDNA arrays within the nucleolus. When the initial transgene copy number exceeds cellular tolerance, endogenous rDNA quality control pathways are activated, driving homologous recombination (HR)-mediated rDNA rearrangement to reduce transgene copy numbers to a tolerable level. This rDNA quality control mechanism is conserved across nearly all cell types tested in this study; notably, the driving force for rDNA recombination is likely derived from endogenous cellular stress² rather than the endonuclease activity of R2 retrotransposons themselves. Post-integration, cell-type-specific regulatory layers further modulate transgene expression. In primary human immune cells (T cells and NK cells), integrated transgenes are subject to robust transcriptional or translational repression in the native immune cellular environment. This repression can be reversed by either ectopic expression of the MERS-CoV ORF4a (Vp4) protein or cellular reactivation stimulation. Additionally, treatment with the cell cycle inhibitor Palbociclib effectively attenuates the loss of transgene expression signals in T cells during long-term culture. In HEK293T cells and non-immune primary cells, persistent and stable transgene expression can be achieved.



Supplementary Fig. 7 | Investigation of gene integration mediated by engineered R2SPs retrotransposons *in vivo*

a–c, *In vivo* integration of the Nano luciferase transgene in mice via LNP-mediated intramuscular delivery of the R2 tool. n = 3 mice. **(a)** Luciferase activity; **(b)** Genotyping analysis; **(c)** Validation of correct integration sequences via Sanger sequencing.

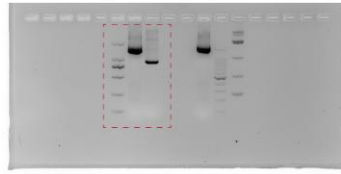
Unless otherwise stated, data are presented as mean $\pm$ SD.



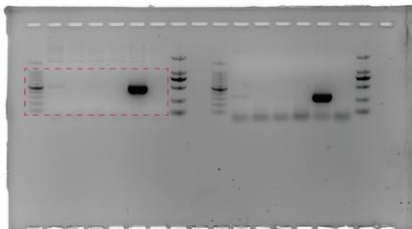
Supplementary Fig. 8 | Flow cytometry gating strategies for distinct cell types used in this study

a, HEK293T cells; **b**, primary human T cells (corresponding to Fig. 2e); **c**, primary human T cells (for other T cell-related figures in this study); **d**, primary human NK cells; **e**, adult primary human skeletal muscle (HSKM) myoblasts; **f**, primary human foreskin fibroblasts.

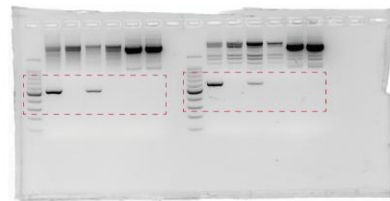
Supplementary Fig. 2a



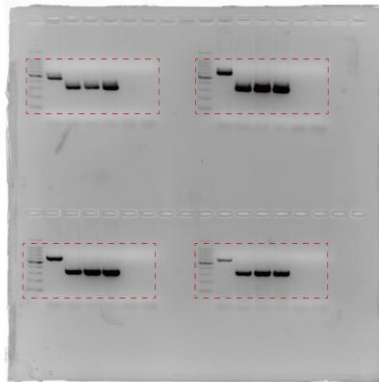
Supplementary Fig. 2c



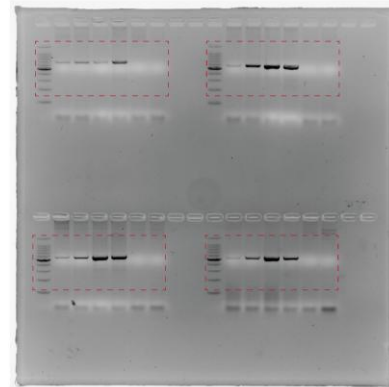
Supplementary Fig. 2g



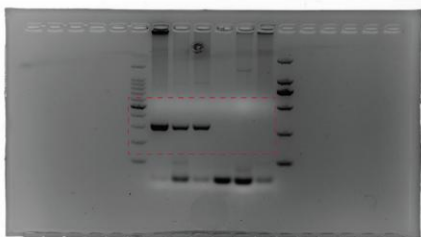
Supplementary Fig. 4b



Supplementary Fig. 4b



Supplementary Fig. 7b



Supplementary Fig. 9 | Unprocessed source gel images in Supplementary Figures

Reference

- 1 Chen, Y. *et al.* All-RNA-mediated targeted gene integration in mammalian cells with rationally engineered R2 retrotransposons. *Cell* (2024). <https://doi.org/10.1016/j.cell.2024.06.020>
- 2 Warmerdam, D. O. & Wolthuis, R. M. F. Keeping ribosomal DNA intact: a repeating challenge. *Chromosome Res* **27**, 57-72 (2019). <https://doi.org/10.1007/s10577-018-9594-z>