nature communications

Article

https://doi.org/10.1038/s41467-026-74164-z

Optimization of IS621 recombinase/bridge RNA-directed recombination for precise insertion of large DNA fragments in human cells

Received: 24 September 2025

Accepted: 27 May 2026

Published online: 06 June 2026

Check for updates

Junfan Guo[1,2,9], Ziguo Song[3,4,9], Wuke Wang[5,9], Shisheng Huang[6], Yu Zhang[7], Guanglei Li[7], Xingxu Huang[1,5], Xiaolong Wang[3,4] ✉, Jianghuai Liu[8] ✉ & Lina Wu[2] ✉

Therapeutic integration of large DNA fragments into precise genomic loci of human cells remains challenging. The IS621 recombinase, derived from the IS110 family, utilizes a bridge RNA (bRNA) containing an independently programmable target-binding loop (TBL) and donor-binding loop (DBL) to mediate sequence-specific recombination between target and donor DNA in prokaryotes. To expand its application to human cells, we rationally engineer the IS621 protein and its cognate bRNA. This ultimately leads to the development of enIS621-tebRNA, which is capable of driving site-specific, scarless insertion of large DNA fragments in human cells. We systematically evaluate the insertion efficiency of this system across multiple cell types, achieving integration rates up to 27.75% for kilobase-scale DNA cargos. Furthermore, we successfully achieve functional insertion of a CD19 chimeric antigen receptor (CAR) at relevant loci for T cell therapy, and establish site-specific integration of the gene encoding human Factor IX with implications for hemophilia B gene therapy. These results highlight the potential of the engineered enIS621-tebRNA system as a versatile platform for precise insertion of large DNA fragments in human cells.

The ability to precisely manipulate the human genome for therapeutic purposes has long been a major aspiration in biomedical research. The CRISPR-Cas system has been widely adopted for diverse genome editing applications due to its programmability and high efficiency. While base editors[1,2] and prime editors[3] have enabled precise modification of single or a few nucleotides, the precise insertion of DNA fragments larger than 1 kilobase at defined genomic loci remains a significant challenge. Classical homology-directed repair (HDR) approaches depend on cell-cycle stage and show markedly reduced efficiency with increasing cargo size[4–6]. Homology-independent

[1]School of Life Science and Technology, ShanghaiTech University, Shanghai, China. [2]State Key Laboratory of Microbial Technology, School of Food Science and Pharmaceutical Engineering, Nanjing Normal University, Nanjing, China. [3]Hainan Institute of Northwest A&F University, Sanya, China. [4]College of Animal Science and Technology, Northwest A&F University, Yangling, China. [5]The Key Laboratory of Pancreatic Diseases of Zhejiang Province, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China. [6]The Second People's Hospital of Changzhou, The Third Affiliated Hospital of Nanjing Medical University, Changzhou Medical Center, Nanjing Medical University, Changzhou, Jiangsu, China. [7]Shanghai-MOST Key Laboratory of Health and Disease Genomics, NHC Key Lab of Reproduction Regulation, Shanghai Institute for Biomedical and Pharmaceutical Technologies, Shanghai, China. [8]State Key Laboratory of Pharmaceutical Biotechnology, Medical School of Nanjing University, Nanjing, China. [9]These authors contributed equally: Junfan Guo, Ziguo Song, Wuke Wang. ✉e-mail: xiaolongwang@nwafu.edu.cn; liujianghuai@nju.edu.cn; wuln@njnu.edu.cn

targeted integration (HITI)[7], based on non-homologous end joining (NHEJ), permits editing in both dividing and non-dividing cells[8], but is associated with safety concerns such as chromosomal translocations, large deletions, and p53 activation[9,10]. Several new strategies, including dCas9-SSAP fusions[11], prime editing-based strategies like PASTE[12], twinPE[13], and AE[14], as well as CRISPR-associated transposases (CASTs)[15,16], have advanced in reducing double-strand breaks or reconfiguring recombination pathways. Nevertheless, these systems often suffer from high complexity or leave behind target site duplications (TSDs), which limit their utility. Alternative transposon-based systems, such as Sleeping Beauty[17] and piggyBac[18], enable large DNA integration, while they lack precise control over insertion sites/copy numbers, are not scarless, and often introduce unwanted terminal transposase-recognition sequences.

Recent studies have shown that IS621[19,20], a member of the IS110 family, encodes a 326-amino acid recombinase with an N-terminal DEDD-type RuvC-like domain and a C-terminal Tnp domain containing a highly conserved serine residue. In the excised, circular form of the transposable element, the non-coding regions flanking the transposase are transcribed into a 177-nucleotide (nt) non-coding RNA. This guide-acting RNA is termed the bridge RNA (bRNA), which mediates dual recognition of donor and target DNA through two discrete loops. The loops base-pair with sequences adjacent to a central CT dinucleotide core in the respective DNA substrates, guiding the assembly of a synaptic complex and enabling recombination via a Holliday junction intermediate in prokaryotes[19–21]. Importantly, the sequences for both loops can be flexibly programmed to mediate target site- and donor-specific recombination in bacteria. Therefore, the bRNA-guided recombination mechanism offers a compact architecture with strong programmability. Compared with other transposase-dependent recombination tools, the IS621-derived system does not require pre-installed recognition sites or complex intermediate steps. Together, these positive attributes have made the IS621-bRNA recombination system a promising platform for precise, large-fragment DNA insertion for potential application in human cells.

Genome editing in cells from higher organisms necessitates high activity of the tools. To enable the applicability of IS621-bRNA in human cells, we perform rational engineering of both the recombinase and the bRNA components. Systematic single-residue mutagenesis and combinatorial screening ultimately led us to establish an enhanced triple mutant variant, enIS621 (A119R/H138R/V293Q), which presents evident activities in human cells. Meanwhile, we developed three alternative bRNA formats (tbRNA/ebRNA/tebRNA), among which tebRNA exhibited the greatest enhancement in editing performance. In combination, the optimized format is named enIS621-tebRNA. We systematically evaluate insertion efficiency, precision, cell-cycle dependence, and genome-wide safety of the engineered system at endogenous loci of various human cell types, including HEK293T, HCT116, RPE-1, Jurkat, and Huh-7 cells. Furthermore, we successfully inserted the genes encoding functional CD19 chimeric antigen receptor (CAR) and human Factor IX at therapeutically relevant loci, demonstrating the potential of the platform for cell and gene therapy applications.

Results

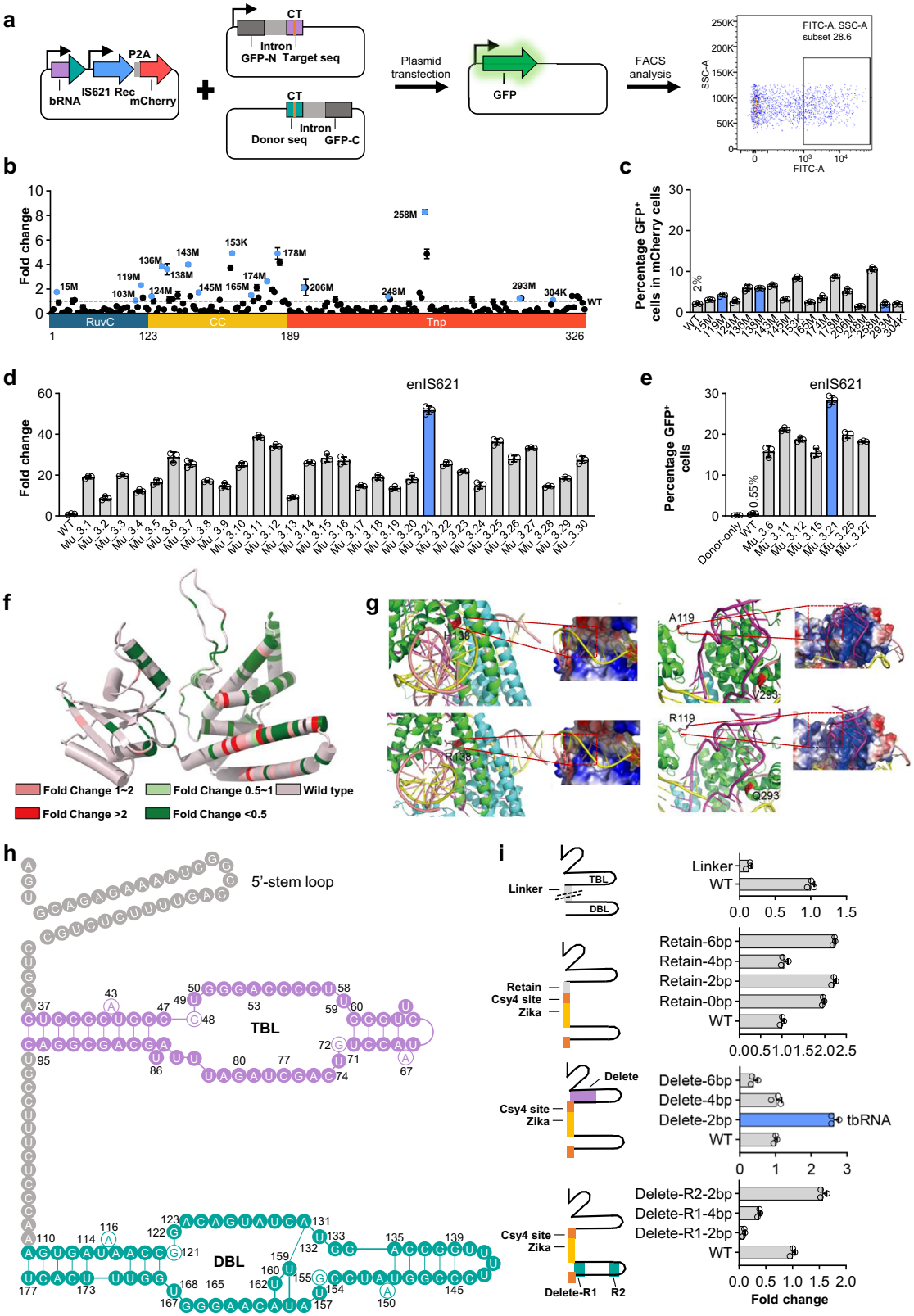
Engineering IS621 recombinase and bRNA for precise insertion at endogenous loci in human cells

Given the evident transposition activity of IS621 in prokaryotic systems (schematic illustration in Supplementary Fig. 1a)[19], we sought to further adapt this recombinase and its programmable bRNA for site-specific genome editing in human cells. To sensitively monitor IS621-mediated insertion events, we developed a fluorescence-based reporter system comprising three constructs: (1) an expression vector encoding the IS621 recombinase, bRNA, and mCherry; (2) a construct harboring the N-terminal half of GFP (GFP-N) and the 5′ half of an intron followed by the target sequence; and (3) a construct containing the donor sequence followed by the 3′ half of the intron and the C-terminal half of GFP (GFP-C)[22,23]. These three plasmids were co-transfected into HEK293T cells. The sequence design for the bRNA (with 14-bp recognition) was according to previous publications and the associated online tool[19,20]. Upon successful recombination of the reporter construct, the full-length GFP would be expressed and produce the fluorescence signal to sensitively indicate IS621 activity (Fig. 1a–e, Supplementary Fig. 1b).

Notably, minimal recombination signals were observed in HEK293T cells when the constructs of the basic IS621-bRNA formats were transfected, suggesting insufficient activities from either the recombinase or bRNA components (see Fig. 1c, “WT”). Therefore, we sought to explore protein engineering to enhance IS621 activity in human cells. To this end, we employed Discovery Studio software to thoroughly screen each residue and predict the mutational impact. Single amino acid substitutions in these residues may enhance binding affinity to either DNA or bRNA (represented by ΔΔG, Supplementary Fig. 2). These computationally identified sites were cross-referenced with sequence alignments of IS621 homologs to identify non-conserved positions (Supplementary Data 1), which were then prioritized for experimental validation. A total of 102 candidate residues was next selected for saturation mutagenesis. For each site, two libraries were generated using the KNM and MNB codon schemes, yielding a total of 204 mini-libraries. These mini-libraries were then screened in HEK293T cells using the GFP reporter system via flow cytometry (Fig. 1b). From this initial screen, we chose 16 mini-libraries that represented the majority of potential hits (blue points in Fig. 1b) showing activities above the basal levels (the WT IS621 group). In a validation experiment, the activity patterns of the 16 mini-libraries were mostly confirmed (Fig. 1c). To pinpoint the specific amino acid substitutions underlying these improvements, we prepared individual mutants from the hit mini-libraries and performed an additional screen. Indeed, one or more variants with apparently enhanced activities could be observed within the series corresponding to each mini-library (Supplementary Fig. 3). For respective positions, the evident hits included E15L, A119R, P124H, V136R, H138R, D143K, H145R, E153Y, A165P, L174K, E178R, E206P, S258R, V293Q, and Q304A (Supplementary Fig. 3). Given that IS621 comprises three structural domains, we chose the best-performing mutations from each domain to construct another series of 30 combinatorial variants (Fig. 1d, annotations in Source Data). Among these, the A119R/H138R/V293Q triple mutant exhibited the most robust improvement in recombination efficiency (Fig. 1d, e). Additional combinations of single mutations did not yield further improvements (Supplementary Fig. 4). We therefore defined the A119R/H138R/V293Q combination (51.89-fold over WT IS621) as the final engineered version, hereafter referred to as enIS621 (engineered IS621).

Collectively, these results established a greatly enhanced IS621 variant capable of driving bridge recombination in human cells. To explore the molecular basis for its improved activity, we mapped the results from the mini-library screen onto the cryo-EM structure of the IS621 transposition complex (PDB: 8WT6) (Fig. 1f). Structural modeling suggests that the coiled-coiled domain-located H138R is situated near the backbone of the duplexed portion of DNA, potentially causing further stabilization of the synaptic complex. Moreover, A119R is located near a position of bRNA strand adjacent to the guiding sequence, possibly strengthening the IS621 association with bRNA (Fig. 1g). On the other hand, given that V293Q variant behaved quite differently from V293Q/K variants (see Supplementary Fig. 3), we speculate that the V293Q effect is independent of direct nucleic acid-binding. We next investigated the characteristics of the selected IS621 variants to complex with the nucleic acid components. Target DNA (tDNA) and donor DNA (dDNA) were biotinylated and separately immobilized onto streptavidin-magnetic beads, which were

https://doi.org/10.1038/s41467-026-74164-z



subsequently used to pull down IS621 variants in recombinase/bRNA-expressing cell lysates (Supplementary Fig. 5a). Western blot analyses of the captured fraction revealed enhanced binding of IS621-H138R and enIS621 to both target DNA and donor DNA (Supplementary Fig. 5b, c). In parallel, biotinylated bRNA was used to capture IS621 variants from recombinase-expressing cell lysates (Supplementary Fig. 5d). The results showed that IS621-A119R and enIS621 displayed increased binding to bRNA, compared with wild-type IS621 (Supplementary Fig. 5e, f). These results suggest that improved binding of enIS621 to substrate DNA and to the bRNA in the recombination synaptic complex contributes its enhanced activities.

To further enhance insertion efficiency, we then focused on engineering the bRNA. Since the IS621 synaptic complex is composed of two recombinase dimers, each engaging one bRNA molecule[20], we

Fig. 1 | Engineering of IS621 recombinase and bRNA to induce large-fragment DNA insertion at endogenous loci in human cells. a Schematic of the fluorescence reporter assay to assess insertion efficiency of IS621 recombinase variants in HEK293T cells. The three-plasmid system contains constructs expressing bRNA, IS621 variants and mCherry; the GFP-N-intron target sequence; and the intron-GFP-C donor sequence. IS621 variants carrying individual amino acid substitutions to KNB or MNB libraries. Each dot represents the relative insertion efficiency of a single substitution normalized to WT-IS621 control (dashed line), with blue dots indicating activity-enhancing substitutions (**b**). The GFP$^+$ patterns (% in mCherry$^+$ cells) for a positive subset of mini-libraries were largely validated. Blue bars denote residues later modified in the enIS621 (**c**). Activities of triple-substitution mutants tested using the reporter system. Relative insertion efficiency of all thirty triple-substitution mutants normalized to WT-IS621 (**d**). Insertion efficiencies of selected active triple-substitution mutants were further compared. WT and mutant efficiencies are shown as the proportion of GFP$^+$ cells within the mCherry$^+$ population, whereas the donor-only group is shown as the proportion of GFP$^+$ in all cells (0.18%) (**e**). Blue bar indicates enIS621. **f** Mapping of the hit (light and dark red) and selected negative (light and dark green) mini-libraries onto the IS621 recombinase structure (PDB: 8WT6). Colors indicate relative activity compared with WT. **g** Structural diagrams highlight the interactions of enIS621 recombinase (PDB:8WT6) with DNA and bRNA. Ribbon diagrams, with H/R138 or A/R119 residues highlighted in red, are paired with charge distribution maps. Left panels suggest potential H138R-DNA interactions; right panels suggest potential A119R-bRNA interactions. **h** Schematic of IS621 bRNA. TBL, target-binding loop; DBL, donor-binding loop. **i** Relative insertion efficiencies of IS621-guided by engineered bRNAs. The bRNA engineering schemes are shown on the left and variant efficiencies on the right. The Csy4 and xrRNA elements are shown in orange and yellow, respectively, and truncated stem regions near TBL and DBL are shaded purple and blue, respectively in the quantitation graph. The tbRNA group is indicated. Data are presented as means ± s.d., $n = 3$ independent biological replicates. Source data are provided as a Source Data file.

hypothesized that splitting the bRNA into distinct TBL and DBL components could reduce structural constraints and therefore enhance editing activity (Fig. 1h). However, direct use of two RNA components, which respectively contained TBL and DBL but were devoid of the linker region, resulted in a significant decrease in transposition efficiency (Fig. 1i, top panel). This decrease was likely due to structural instability and rapid degradation of the split RNA fragments. To address this possible limitation, we introduced a Csy4 recognition site between the TBL and DBL to enable post-transcriptional intra-molecular cleavage, with the Csy4 element also providing a degradation-resistant secondary structure at the 3′ end[24]. We additionally inserted a Zika-derived xrRNA element on the 5′ side of the DBL, which could lead to stabilization of the DBL component following Csy4 cleavage[25–27]. Hence, the resultant construct featured incorporation of these cleavage and protection motifs in place of the original linker sequence ("retain 0 bp"). Indeed, such a configuration (together with Csy4) led to around 2-fold increase of recombination events compared with the basal level (WT IS621-bRNA) in transfection experiments (Fig. 1i, second panel). Furthermore, retaining 0-, 2-, or 6-nt linker sequences downstream of TBL in such modified bRNA did not appear to cause changes in recombination rates (Fig. 1i, second panel). On the basis of the active, cleavage-dependent format (Csy4/xrRNA), we further assessed the impact of the loop-adjacent stem region on recombination activity. We introduced incremental 2-bp truncations in the stem region along G37 ~ C42 toward the TBL. In addition, on either side of the DBL, small stem truncations were introduced along A110 ~ G114 ("R1", 2- or 4-bp), and at G133-G134 ("R2", 2-bp), respectively. These deletional variants exhibited differential activities relative to the original bRNA (WT) format. Overall, the Csy4/xrRNA configuration with a 2-bp deletion at G37-U38 (paired with A94-C95) near TBL exhibited the highest insertion efficiency among all tested bRNA variants (Fig. 1i, 3rd and 4th panels). This optimized, two-part bRNA, which was 2.64-fold more efficient than bRNA, was named tbRNA.

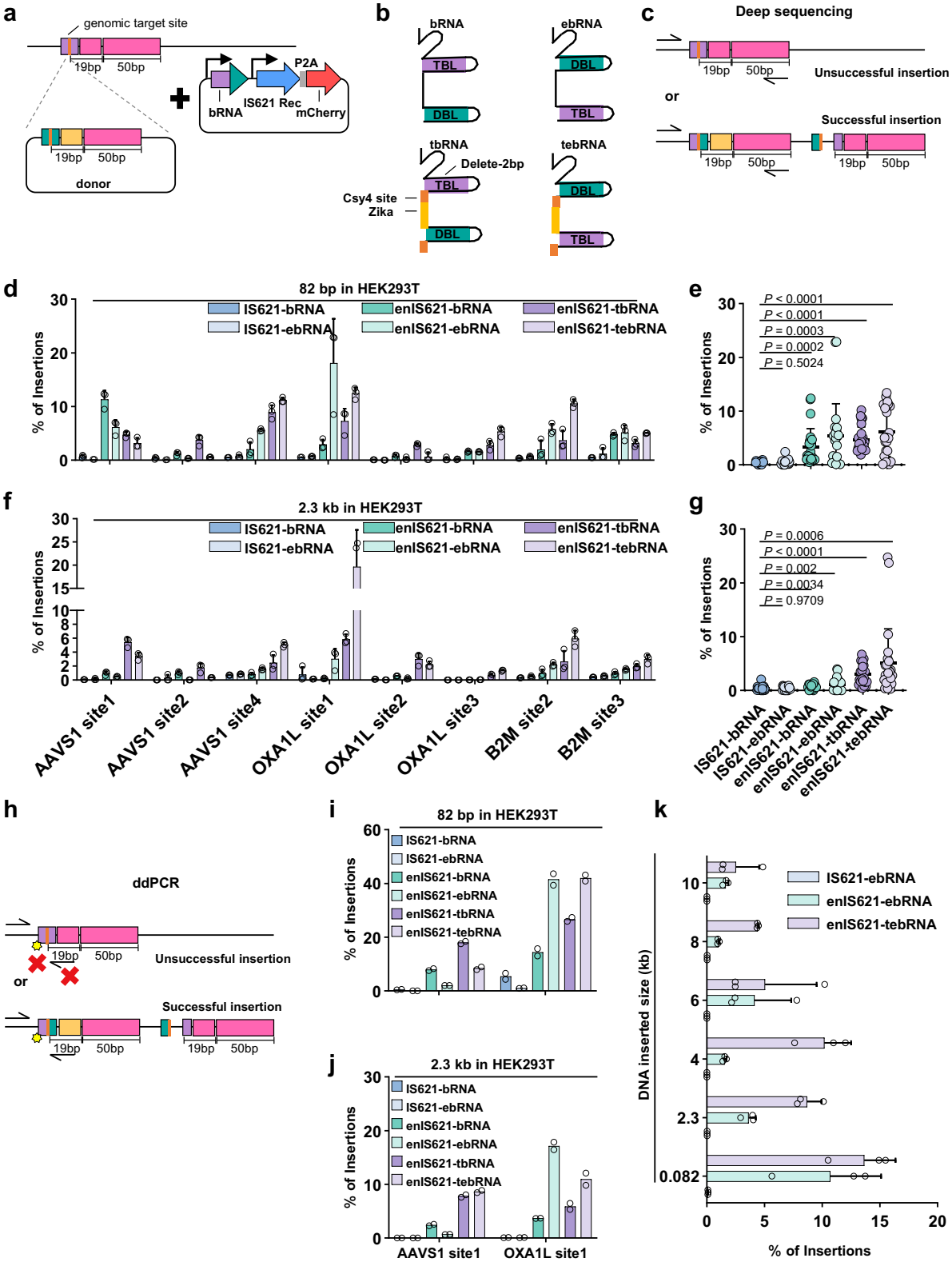
Application of enIS621 for precise DNA insertion at endogenous loci in various human cells

We next evaluated the insertion efficiency of our engineered IS621 recombinase and bRNA system at endogenous genomic loci of human cells (Fig. 2a). Given the strong programmability of bRNA, either the TBL or DBL can be designed to recognize and bind to endogenous genomic sequences. Interestingly, from our target or donor DNA pull-down experiments, IS621-bRNA appeared to exhibit a higher binding affinity for dDNA (Supplementary Fig. 5b, c). In consequence, we also explored an alternative design of bRNA in which the base-pairing properties of the binding loops were exchanged. Herein, we designated a bRNA that instead use the DBL to recognize the target locus and the TBL to bind the donor sequence as ebRNA (the exchange design variant of bRNA). Along this line, a similar alternative version of tbRNA was referred to as tebRNA (two-part and exchange design variant of bRNA) (Fig. 2b).

We selected eight endogenous loci in the HEK293T genome to systematically assess insertion efficiency. To facilitate downstream analyses, a 50-bp sequence identical to the region downstream of the endogenous CT core was placed 19 bp downstream of the donor CT core. This donor design would enable amplification of WT and recombined alleles with a common primer pair (Fig. 2c). We co-transfected the cells with IS621 or enIS621, various formats of bRNA, and the donor. Deep sequencing showed that the unmodified IS621 paired with wild-type bRNA or with ebRNA was ineffective in mediating donor insertions. For an 82-bp circular DNA donor, the insertion efficiencies (average ± SD) by bRNA and ebRNA-guided IS621 were 0.41% ± 0.23% and 0.49% ± 0.52%, respectively, across eight loci. For a 2.3 kb DNA donor, the insertion efficiencies by the two IS621 groups were 0.31% ± 0.44% and 0.31% ± 0.32%, respectively (Fig. 2d–g). In comparison, all enIS621 groups (guided by bRNA, ebRNA, tbRNA, or tebRNA) demonstrated significantly enhanced performances. For the 82-bp payload, insertion efficiencies (average ± SD) by enIS621 were 3.33% ± 3.42% (bRNA), 5.35% ± 6.01% (ebRNA), 4.66% ± 2.42% (tbRNA), and 6.12% ± 4.57% (tebRNA). enIS621 also achieved evidently higher efficiencies over IS621 for integration of the 2.3 kb donor, reaching levels of 0.69% ± 0.42% (bRNA), 1.13% ± 1.16% (ebRNA), 2.98% ± 1.86% (tbRNA), and 5.12% ± 6.34% (tebRNA). Consistent with the reporter results (see Fig. 1i), the two-part bRNAs generally showed greater activities than their single-entity counterparts. For all tested groups, enIS621-tebRNA overall presented the highest efficiencies to integrate both donors (Fig. 2e, g). Transcriptional activators have previously been shown to improve genome editing outcomes by increasing chromatin accessibility[15,28]. However, fusion of the transcriptional activator VPR to enIS621 did not boost insertion efficiency (Supplementary Fig. 6a, b). To validate our insertion quantifications by orthogonal approaches, we applied digital droplet PCR (ddPCR) (Fig. 2h-j, Supplementary Fig. 6c, d) and absolute qPCR (Supplementary Fig. 6e, f) at two representative loci, *AAVS1* site1 and *OXA1L* site1. These independent quantitative analyses largely confirmed the deep sequencing results for these sites (see Fig. 2d, f, and Supplementary Fig. 6a, b for references). We also evaluated the length limit of enIS621-mediated insertions. The integration rates decreased with increased donor size. Nevertheless, with the largest, 10 kb donor applied in the test, we still observed integration efficiencies of 1.64% and 2.53% in response to enIS621-ebRNA and enIS621-tebRNA, respectively (Fig. 2k).

In addition to efficiency, we further assessed the precision and integrity of DNA integration mediated by enIS621. The allele types and frequencies at the junctions between endogenous target sites and exogenous donor DNA were inferred from deep sequencing data (Supplementary Fig. 6g). Across the outcomes corresponding to all four bRNA formats, the proportion of precise alleles within total

https://doi.org/10.1038/s41467-026-74164-z



integrated alleles exceeded 91%, indicating that enIS621-bRNA enabled highly accurate end-to-end recombination. Subsequently, we assessed the structural integrity of the integration products by long-read sequencing. To this end, two partially overlapping cross-junction amplicons (Amplicon 1 and Amplicon 2) were prepared to span the integrated payload and the flanking genomic sequences (Supplementary Fig. 6h). The amplified products were subjected to PacBio long-read sequencing. Analysis of the resulting reads from all bRNA groups revealed no detectable structural abnormalities or high-frequency mutations within the donor sequence or at the donor-genome junctions (Supplementary Fig. 6i). The distribution of indels across the integration region was comparable across groups corresponding to different bRNA formats. These results indicate that enIS621-mediated recombination proceeds with high precision and

Fig. 2 | enIS621-mediated insertion editing at endogenous loci in HEK293T cells using various bRNA formats. a Schematic illustration of IS621 recombinase-mediated insertion at endogenous genomic loci. To simplify quantitative analyses, the sequence downstream of the recognition site in the donor is engineered to feature a unique 19-bp sequence (for quantitation of donor insertion) followed by a 50-bp sequence identical to the genomic sequence downstream of the integration site (for amplification of integrated allele). **b** Schematic of WT-bRNA and three engineered bRNA variants. **c**–**g** Deep sequencing analysis of insertion efficiency for an 82 bp donor or 2.3 kb donor at endogenous loci mediated by IS621 or enIS621 recombinases with various bRNA formats. The PCR scheme for amplification of the WT and the knock-in allele is shown in (**c**). A 50-bp sequence downstream of the endogenous target CT core (19 bp) was appended similarly downstream of the donor CT core, allowing amplification with a common primer pair. The data for insertion efficiency comparisons across eight endogenous loci are shown in (**d**, **f**). The dot plot summarizing results from all tested sites are shown in (**e**, **g**). **h**–**j** ddPCR analysis of insertion efficiency at selected endogenous loci by IS621 or enIS621 recombinases with various bRNA formats. **h** The PCR scheme is illustrated. The forward primer was positioned upstream of the endogenous insertion site and the reverse primer within the donor-specific sequence, ensuring amplification of integrated sequences. The small yellow symbol indicates the position of the TaqMan probe. The scheme also highlights the inability of WT allele to produce signal. The mean efficiencies of IS621 and enIS621 (with different bRNA formats) for integrating an 82 bp donor (**i**) and a 2.3 kb donor (**j**) are shown, $n = 2$ biological replicates. **k** Comparison of insertion efficiencies for donors of different lengths mediated by IS621-bRNA, enIS621-ebRNA, and enIS621-tebRNA. The integration efficiencies were determined by deep-sequencing-dependent quantitation. Deep sequencing data in this figure are presented as means ± s.d., $n = 3$ independent biological replicates. *P* values were calculated by Two-tailed unpaired Student's *t*-test. Source data are provided as a Source Data file.

preserves the structural integrity of the target site in mammalian cells.

To evaluate the generalizability of enIS621-mediated large-size DNA insertion across human cell types, we extended our investigation to HCT116 cells. As expected, enIS621-tebRNA exhibited the highest insertion efficiency among the tested bRNA formats (Supplementary Fig. 7a–d). Specifically, when guiding the integration of an 82 bp exogenous DNA donor at five endogenous genomic loci, the insertion efficiency (average ± SD) reached 7.85% ±8.98% (Supplementary Fig. 7a, b). Notably, when directing the insertion of a longer 2.3-kb donor sequence at the same sites, the efficiency was 4.09% ± 5.23% (Supplementary Fig. 7c, d). These results support the robustness and applicability of the enIS621 system for programmable DNA integration across distinct human cell lines.

In addition, we examined the effect of cell cycle status on enIS621-mediated insertion. To this end, we employed RPE1 cells, a diploid immortalized cell line commonly used for investigation of cell cycle regulation and for studies of genome editing[14,29]. RPE1 cells were arrested in S-phase by aphidicolin treatment, which was confirmed by inhibition of EdU incorporation (Supplementary Fig. 8a, b). Following ±aphidicolin pre-treatment, cells were transfected with enIS621-tebRNA and the donor. The aphidicolin was removed 3 days following transfection, to avoid extensive cellular toxicity. One day later, the cells were harvested for analyses of donor insertion. Across four endogenous loci, the insertion efficiency (average ± SD) was 4.34% ± 2.84% in untreated cells, and 8.00% ± 3.55% in aphidicolin-treated cells (Supplementary Fig. 8c). These results indicate that enIS621-mediated donor insertion does not require active cell cycle progression.

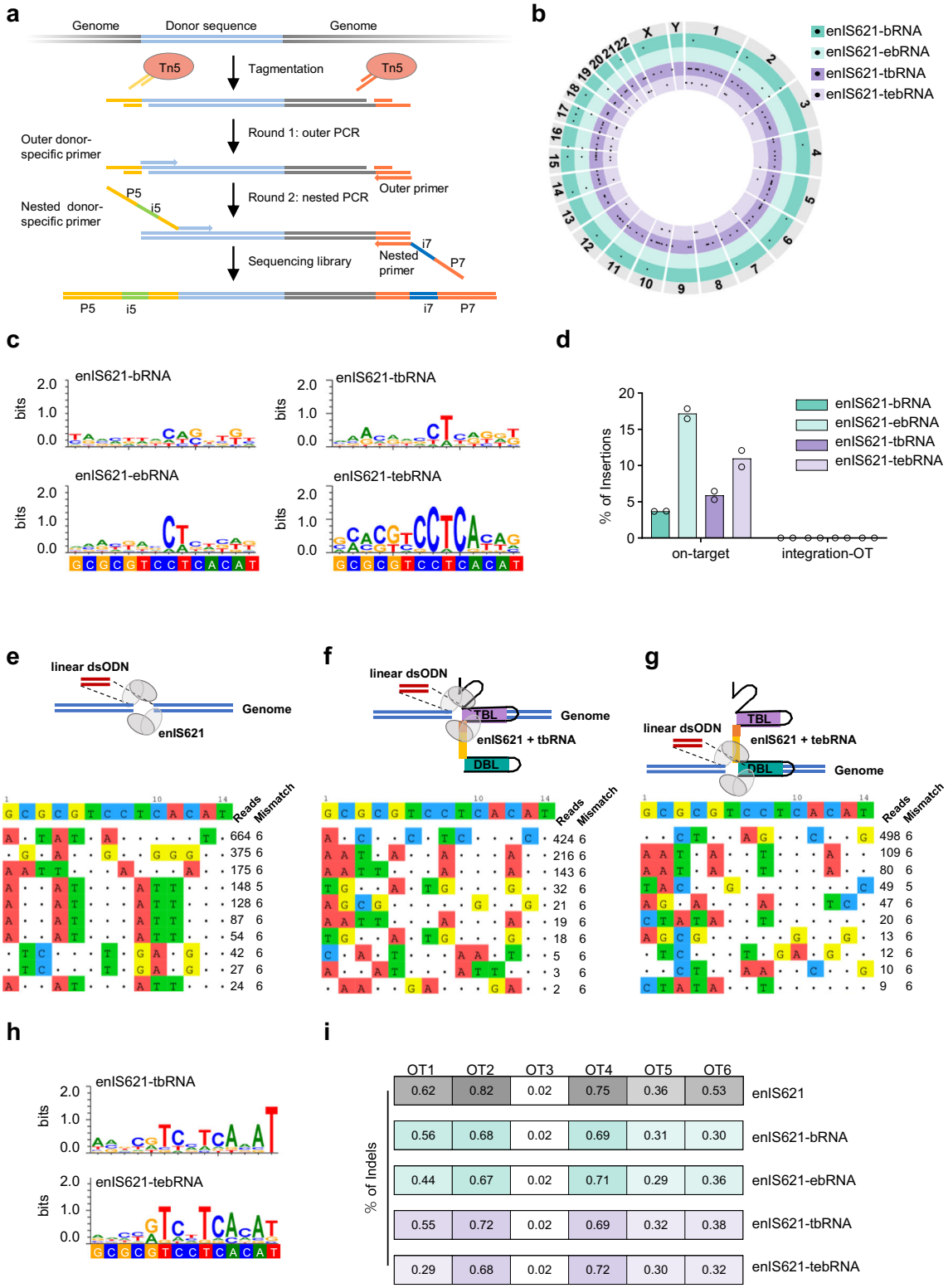
Given that another member of the IS110 family, ISCro4[30,31], has been reported to enable site-specific recombination in human cells, we next performed a direct comparison of its insertion capability with that of enIS621. Insertion efficiencies were systematically evaluated at five endogenous genomic loci. For the 82 bp donor sequence, enIS621-tebRNA mediated an insertion efficiency of 14.71% ± 12.44% (average ± SD). In parallel, the ISCro4 system was tested using the configuration reported to yield the highest activity in the literature, namely the engineered enISCro4 recombinase combined with a split bRNA in which the donor-binding loop targets the genomic sites[31]. Under these conditions, ISCro4 achieved an insertion efficiency of 12.49% ± 11.29% (Supplementary Fig. 9a, b). For a 2.3 kb donor sequence, enIS621-tebRNA achieved an insertion efficiency of 6.57% ± 4.81%, whereas the ISCro4 system yielded an efficiency of 3.03% ± 1.53% (Supplementary Fig. 9c, d). These results suggest that, while enIS621-tebRNA and ISCro4 exhibit comparable insertion efficiencies for short DNA payloads, enIS621-tebRNA appears to present higher activities for integration of larger DNA fragments. Notably, enIS621 and ISCro4 showed some correlations in activity patterns across different sites, consistent with their similarity in editing mechanisms[19,20,30,31].

We also considered to compare the performances of enIS621-tebRNA to another very recent CRISPR-associated transposase tool, the evoCAST system[16], under largely identical experimental settings (Supplementary Fig. 10a). Specifically, we employed evoCAST to insert a 2.3 kb exogenous DNA fragment at four endogenous genomic loci also under test above (see Supplementary Fig. 9c). The stoichiometry of all evoCAST plasmids was used according to the optimized format[16], as the total amount per transfection was set up to contain a similar input of donor DNA plasmid (600 ng) as in the enIS621 experiment (500 ng). We found that across all four loci, the insertion efficiencies achieved by evoCAST were below 1% (Supplementary Fig. 10b), which is substantially lower than the efficiencies observed with enIS621-tebRNA (see Supplementary Fig. 9c). Unlike the ISCro4 results, the cross-site pattern of evoCAST activities did not correlate with those of enIS621, most likely determined by the distinct editing mechanisms for IS110 recombinases and CRISPR-associated transposases[15,16,19,20]. In addition, the evoCAST-induced integration products inevitably contain the transposon's right/left end sequences (RE/LE) and the target site duplications (TSDs) (Supplementary Fig. 10a), despite featuring precise junction sequences (Supplementary Fig. 10c, measured base substitutions <1.2%). It is also worth noting that the evoCAST system depends on the co-delivery of seven components, including Cas6, Cas7, Cas8, TnsAB, TnsC, TniQ, and crRNA-donor plasmids, which might necessitate some efforts of user optimization. Together, the comparisons with evoCAST underline the favorable activity, minimal scar formation and compositional simplicity as important positive attributes of enIS621-tebRNA.

Genome-wide off-target activity assessment

Considering that IS621-bRNA recognizes the site of integration through bRNA-dependent, ~14 bp base pairing with the target DNA, we systematically evaluated its potential risk in inducing off-target donor integration in the genome. To track the events of donor DNA integration in the genome, we first applied a modified UDiTaS strategy[32,33]. Such a method allows unbiased detection of donor integration sites via tagmentation, target-enrichment, followed by high-throughput sequencing analysis. Cells were co-transfected with enIS621-bRNA, enIS621-ebRNA, enIS621-tbRNA, or enIS621-tebRNA targeting *OXA1L* site1 together with a 2.3 kb circular donor DNA, and two independent biological replicates were taken for each condition (to identify common integration sites). Four days after transfection, fluorescent positive cells were isolated by flow cytometry. Genomic DNA was then extracted, and the sequencing libraries were prepared according to the modified UDiTaS workflow to enrich and sequence the donor-genome junction fragments. The contaminating reads corresponding to the transiently transfected donor plasmid were filtered by removing reads containing a distal plasmid sequence. The resulting reads were next mapped to determine the donor integration sites across the genome (Fig. 3a).

https://doi.org/10.1038/s41467-026-74164-z



Such analysis of genome-wide integration events revealed that different bRNA configurations influenced off-target integration profiles. A total of 42, 23, 139, and 48 off-target integration sites were detected for enIS621-bRNA, enIS621-ebRNA, enIS621-tbRNA, and enIS621-tebRNA, respectively (see Supplementary Data 2). The genomic distribution of integration sites in each condition is presented in Circos plots (Fig. 3b). Motif analysis of the top ten sites ranked by read abundance highlights a frequent CT dinucleotide core, supporting that the captured integration sites contain bona fide off-targets by the recombinase (Fig. 3c). A clear similarity of tebRNA motif to the desirable target site could also be recognized, providing further confidence to this analysis pipeline. Notably, the increased number of off-target integration sites by the two-part bRNA variants (tbRNA and tebRNA) compared with the single-unit bRNA configurations (bRNA and

Fig. 3 | Genome-wide off-target activity assessment. a–d Genome-wide off-target integration analyses. Schematic illustration of the modified UDiTaS workflow used to detect genome-wide donor integration sites. Genomic DNA was first randomly fragmented and tagged using Tn5 transposase. Two rounds of PCR amplification were then performed, respectively pairing outer and nested donor primers with the corresponding adaptor-end primers. This led to enrichment of donor-genome junction fragments suitable for Illumina sequencing (**a**). **b** HEK293T cells were transfected with enIS621-bRNA, enIS621-ebRNA, enIS621-tbRNA, and enIS621-tebRNA targeting the *OXA1L* site1 locus. The genome-wide donor integration events were captured with the workflow above, and the sites were recovered by paired-end Illumina sequencing. Circos plot is used to show the genome-wide distribution patterns of detected donor integration sites. **c** Sequence motifs of the top ten off-target integration sites (according to read abundances) corresponding to various bRNA formats are shown. **d** Samples from enIS621-transfected cells were subjected to ddPCR analyses at a top UDiTaS-nominated off-target integration site. On-target integration frequencies, as also presented in Fig. 2j, are shown for comparison (**d**). Genome-wide off-target cleavage analyses. GUIDE-seq analyses were performed in cells transfected with enIS621 ± tbRNA or tebRNA. The sequence features of GUIDE-seq-captured sites are shown (**e**–**g**) corresponding to no bRNA, tbRNA and tebRNA, respectively). The top ten GUIDE-seq sites in alignment with the target sequence are presented. **h** Sequence motifs of the top ten GUIDE-seq-nominated sites corresponding to tbRNA and tebRNA were determined. **i** Samples from enIS621-transfected cells were subjected to targeted sequencing analyses at a total of 6 GUIDE-seq-nominated sites comprising the respective top-3 sites for tbRNA and tebRNA. The levels of indels (% of total reads) at these sites in control, bRNA, ebRNA, tbRNA and tebRNA samples are shown in a heat map. Data are presented as median values; ddPCR, $n = 2$; deep sequencing $n = 3$ independent biological replicates. Source data are provided as a Source Data file.

ebRNA) points to the decreased targeting specificity upon splitting the TBL and DBL, as suggested by a very recent study on ISCro4[31]. This may represent a potential applicational tradeoff for the more active tbRNA and tebRNA. Interestingly, the substantially reduced numbers of off-target integration sites for ebRNA and tebRNA (23 and 48, *vs* the TBL-target binding configuration in bRNA and tbRNA: 42 and 139, respectively) suggest that DBL-dependent selection of target sites improved integration specificity. Such an observation is also consistent with those from the very recent study on ISCro4[31].

Despite the detection of off-target integration by all bRNA formats, as we did not include UMI in the donor plasmid, our modified UDiTaS protocol might be influenced by PCR biases and therefore might not be quantitative. For quantitation of off-target activities, we performed ddPCR analyses. We chose the on-target site, as well as a high-ranking UDiTaS-nominated off-target site across all experimental groups. The ddPCR results validated on-target integration in all bRNA groups. In contrast, ddPCR analyses at the nominated OT site did not reveal detectable integrations (Fig. 3d). Although the high sensitivity and comprehensiveness of the UDiTaS method are clearly suited to survey the risk of off-target integration, our results above suggest that some of the sites might be false-positives or arising from rare events, especially in transiently transfected cell populations. As it is currently impractical to examine all UDiTaS-nominated sites via ddPCR, we envision that future advances in technologies may be instrumental to definitively assess the off-target integration risk by IS621 recombinases. Overall, our genome-wide analyses of enIS621-associated integration events underscore the ongoing need to improve the current toolset and to explore related systems regarding better integration specificities.

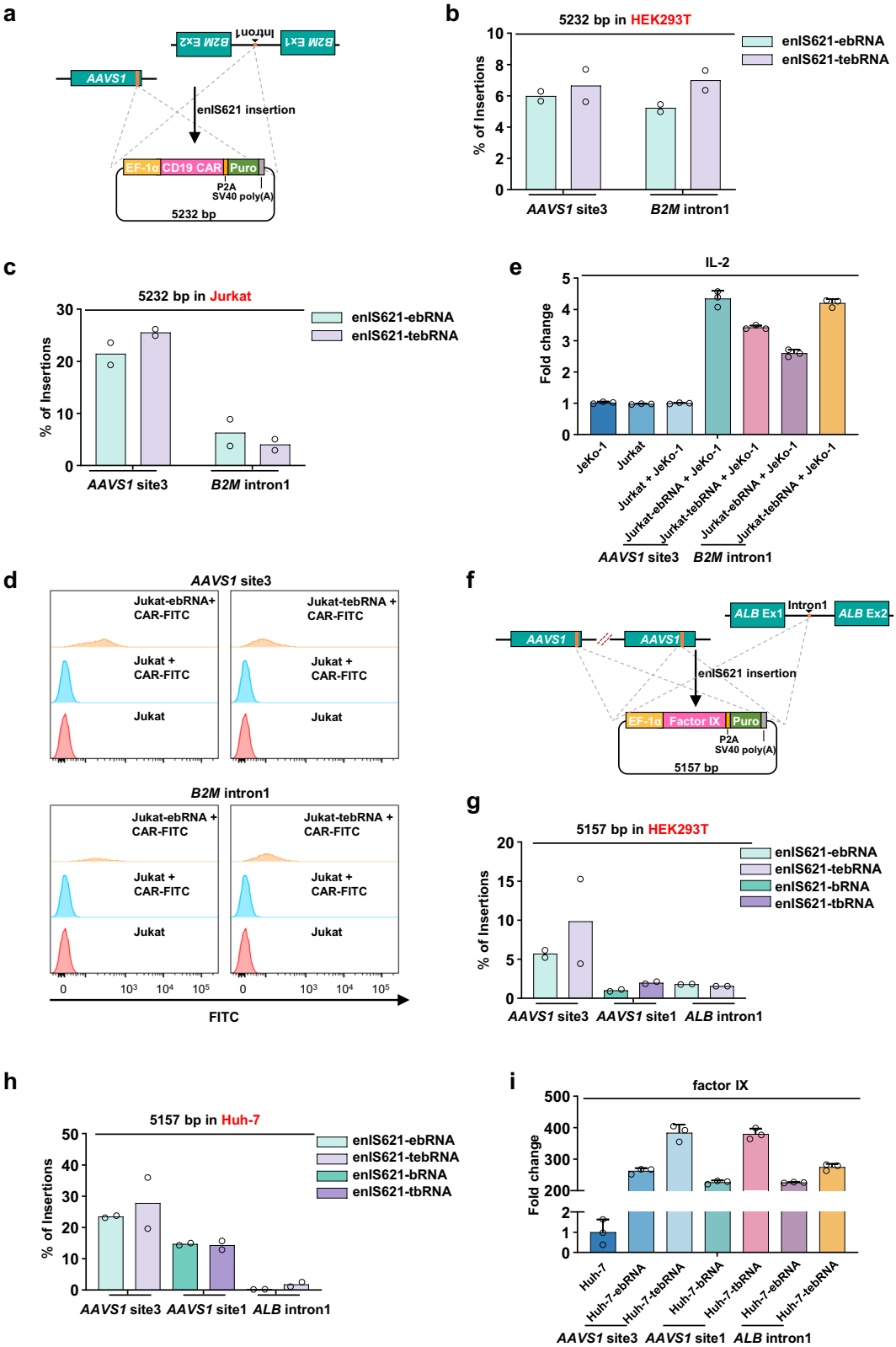
In addition to off-target integration, we further examined whether the IS621 system exhibits potential off-target cleavage activity. To address this question, we performed GUIDE-seq analysis. Cells were electroporated with enIS621 expression plasmids and either of the two higher-activity bRNA configurations (± tbRNA or tebRNA), together with a double-stranded oligodeoxynucleotide tag (dsODN). High-throughput sequencing was then used to identify the sites of dsODN integration (Fig. 3e–g). GUIDE-seq detected DSB signals at only a small number of genomic loci in tbRNA and tebRNA groups. In addition, the general abundance of such enIS621-associated GUIDE-seq reads was in a similar scale as those from the control group (no bRNA) (see Supplementary Data 3). Interestingly, motif analysis of the top-ten sites from the bRNA groups (but not those from the control group) showed some overall similarities to the target DNA sequence, yet without obvious selection of the CT core motif (Fig. 3h). Subsequently, the top three GUIDE-seq-nominated loci from the tbRNA and tebRNA groups were collected (OT1 to OT6) for further testing. Genomic DNA from cells edited with (or without) the four bRNA configurations was amplified by PCR and subjected to deep sequencing analysis. Only background levels of indels could be observed at these sites in samples from all enIS621-bRNA groups (bRNA, ebRNA, tbRNA and tebRNA) (Fig. 3i). Together, although GUIDE-seq captured some weak signal of bRNA-associated off-target cleavage, our deep sequencing results at the candidate sites indicated that the extent of DSB risk by enIS621-bRNA was minimal. This proposition is consistent with the reported DSB-independent mechanism of IS621, even at the target site, where the cleavage of the second target strand only occurs upon completion of the first strand exchange[20].

The enIS621 successfully mediated the precise insertion of cellular and gene therapy-associated coding cassettes

We next evaluated the potential of enIS621 for therapeutic applications involving large-fragment DNA insertion in human cells. Chimeric antigen receptor T cell (CAR-T) therapy represents one of the most transformative technologies in the recent immunotherapy landscape[34,35]. Building on this clinically validated strategy[36,37], we aimed to test enIS621 for precise insertion of a CD19-CAR expression cassette into human cells, thereby exploring its translational potential. Specifically, we utilized circular plasmids encoding CD19-CAR, and employed enIS621-ebRNA and enIS621-tebRNA to induce targeted insertion at two clinically relevant genomic loci harboring a central CT core motif (Fig. 4a). In HEK293T cells, average insertion efficiencies were 5.60% for enIS621-ebRNA and 6.82% for enIS621-tebRNA at the two endogenous loci (Fig. 4b). In Jurkat T cells, enIS621 facilitated efficient insertion, achieving 13.87% for enIS621-ebRNA and 14.76% for enIS621-tebRNA (Fig. 4c). Following three weeks of puromycin selection in Jurkat cells, we performed flow cytometry analysis and validated surface expression of CD19-CAR in the knock-in cells (Fig. 4d, Supplementary Fig. 11a). To assess functional CAR activation, we co-cultured the modified Jurkat cells with CD19-positive JeKo-1 target cells, observing enhanced IL-2 secretion upon stimulation, as measured by ELISA (Fig. 4e). These findings demonstrate that the CD19-CAR sequence is specifically integrated, properly expressed, and responsive to antigen stimulation with an induced cytokine response[38,39]. Notably, junction analysis confirmed precise editing at both 5′ and 3′ boundaries between donor and target sequences, achieving over 98% fidelity in both cell types (Supplementary Fig. 11b).

To further evaluate the therapeutic potential of enIS621 in the context of gene therapy, we chose a target for hemophilia B gene therapy. Hemophilia B is an X-linked bleeding disorder caused by mutations in the coagulation factor IX (*F9*) gene[40,41]. Notably, clinical studies have shown that restoring even as little as 1% of normal factor IX levels can significantly improve coagulation function in patients[42]. To assess the feasibility of precise therapeutic gene insertion, we delivered a circular plasmid encoding human *F9* into two endogenous loci using enIS621-ebRNA and enIS621-tebRNA, and into another locus using enIS621-bRNA and enIS621-tbRNA (Fig. 4f). In HEK293T cells, the insertion efficiencies for ebRNA- and tebRNA-guided

https://doi.org/10.1038/s41467-026-74164-z



enIS621 were respectively 3.75% and 5.71% (average, 2 sites), while those for the bRNA and tbRNA counterparts were 1% and 1.99% (Fig. 4g). In Huh-7 cells (hepatocyte-derived tumor cell line) that are relevant for studying *F9* expression, the insertion efficiencies for ebRNA- and tebRNA-guided enIS621 were respectively 11.86% and 14.8% (average, 2 sites), while those for the bRNA and tbRNA counterparts were 14.69% and 14.31% (Fig. 4h). Following three weeks of further puromycin selection, we assessed factor IX protein expression in edited Huh-7 cells by ELISA and observed a substantial increase in its levels in all groups corresponding to different enIS621-bRNA conditions (Fig. 4i). Furthermore, junction analysis showed precise editing at both 5′ and 3′ boundaries between the target site and the donor sequence, achieving over 94% fidelity in both HEK293T and Huh-7 cells (Supplementary Fig. 11c).

Fig. 4 | enIS621-mediates DNA insertion at therapeutically relevant endogenous genomic loci in multiple human cell types. a–e Insertion of CD19 CAR sequences in HEK293T and Jurkat cells and functional validations. Schematic depiction of CD19 CAR insertion mediated by enIS621-ebRNA and enIS621-tebRNA at *AAVS1* site3 and *B2M* intron 1 (**a**). Digital droplet PCR analysis of CD19 CAR insertion efficiency in HEK293T (**b**) and Jurkat cells (**c**) following transient transfection of the recombination mix. The cells were further selected under puromycin for three weeks. Expression of CAR in Jurkat cells were determined by FACS analyses (**d**). After 24 h co-culture with JeKo-1 cells, IL-2 secretion from different groups of Jurkat cells were measured. The relative levels of IL-2 are shown in (**e**). **f–i** Insertion of *F9* sequences in HEK293T and Huh-7 cells and functional validation. Schematic depiction of *F9* insertion mediated by enIS621-ebRNA, enIS621-tebRNA, enIS621-bRNA, and enIS621-tbRNA at *AAVS1* site1, *AAVS1* site3, and *ALB* intron1 (**f**). Digital droplet PCR analysis of *F9* insertion efficiency in HEK293T (**g**) and Huh-7 cells (**h**) following transient transfection of the recombination mix. The cells were further selected under puromycin for three weeks. Cells were reseeded in 96-well plates for measurements of Factor IX secretion. Seventy-two hours after seeding, the relative Factor IX secretion levels in the WT and knock-in Huh-7 cells were evaluated (**i**). For ddPCR, bars represent mean values (n = 2 biological replicates), with individual data points shown. For ELISA assays, data are presented as means ± s.d., n = 3 independent biological replicates. Source data are provided as a Source Data file.

Discussion

The bridge RNA, composed of fully customizable target-binding and donor-binding loops, enables IS621 recombinase to mediate programmable and site-specific DNA insertion, inversion, and excision at arbitrary loci in prokaryotic genomes[19]. Compared with existing targeted integration strategies, such as prime editing-mediated preinstallation of recombinase recognition sites[43,44], CAST systems that generate target site duplications (TSDs)[45], and transposon systems like Sleeping Beauty and piggyBac with poorly controlled integration sites[46], IS621 offers several distinct advantages. These include a simple system architecture, precise control over insertion site specificity, and scarless integration[19], highlighting its unique potential to evolve into a large-fragment gene insertion tool for cells from higher organisms.

In this study, we rationally engineered the IS621 protein and its bridge RNA to seek a hyperactive variant(s) to sufficiently function in human cells. Our results provided the key demonstration that an optimized, triple residue-substituted IS621 (enIS621) presented evident activities in human cells. We also showed that the bRNA component could be flexibly reconfigured. Indeed, splitting the bRNA's two DNA-binding loops and exchanging their target/donor specificities further enhanced the recombinase activity (the hitherto most active version named "tebRNA"). The enIS621-bRNA variants enabled highly programmable, efficient, cell-cycle-independent, and precise insertion of large DNA fragments into various human cell types. Furthermore, we showcased the feasibility and functional effectiveness of this system in both cellular and gene therapy contexts. It is interesting to note that the current IS621 platform features binding loops of the bRNA with a strong sequence preference for the canonical CT core motif[19,20]. Such a constraint suggests that future development could benefit from AI-assisted design of protein and RNA components, or from mining and reprogramming recombinase candidates derived not only from prokaryotic IS110 elements but also from other prokaryotic and even eukaryotic transposon families[47,48]. This line of research shall ultimately lead to development of recombinases with relaxed sequence constraints, improved specificity, and enhanced integration efficiency, thereby advancing their translational potential in gene therapy.

While our work was in preparation and through the reviewing process, we became aware of two successive reports describing the application of ISCro4, another member of the IS110 family, for genome editing in human cells[30,31]. These studies independently demonstrate that ISCro4 bridge recombinases can mediate programmable genome rearrangements in mammalian cells, yet their research focuses display some differences. Perry et al. devote considerable efforts to optimization of the tool and to its application in large-scale genome rearrangements[31]. Pelea et al. provide cryo-EM-based structural insights and further validate the system's programmability[30]. Herein, our work uncovers the potential of the initially discovered IS621 system in human cells. While we concentrate on the insertion capacity of enIS621 in human cells, its potential in driving other forms of recombination may be anticipated. We also present an in-depth investigation into the compatibility of cargo sizes and targeting accuracies of the inserted sequences as well as their functional outputs, establishing a strong foundation for future development of the technology. Indeed, our data show some correlations in the activity profiles for enIS621 and ISCro4, consistent with their mechanistic similarities. Given the anticipated continuous activity upgrades in IS621 and ISCro4 systems, future comparisons and improvements of their specificities will be highly warranted. Together, the above works by others[30,31] and the present study have advanced the current research on the IS110 family, showcasing the broad application potential of prokaryotic transposon-based platforms for precise genome editing in human cells.

Methods

Plasmid construction

The codon-optimized human expression of the IS621 recombinase gene was synthesized by Tsingke Biotechnology Co., Ltd. (Beijing, China). The U6 promoter and bridge RNA (bRNA) sequence were subsequently introduced to generate an all-in-one plasmid construct. The design for bRNAs (with 14-bp recognition) was aided by the online Bridge RNA Design Tool (https://bridge.hsulab.arcinstitute.org/). Site-directed mutagenesis of IS621, involving the introduction of specific amino acid substitutions, as well as replacement of bRNA sequences targeting different endogenous loci, was performed using the plasmid as a template. PCR amplification was carried out using Phanta Max Super-Fidelity DNA Polymerase (Vazyme), and the resulting amplicons were assembled using the ClonExpress Ultra One Step Cloning Kit V3 (Vazyme) according to the manufacturer's instructions. For short donor DNA circularization, the annealed primers were subjected to ligation overnight. The desired product was harvested by gel purification. Detailed information on endogenous target loci and corresponding bRNA sequences is provided in Supplementary Data 4.

Cell lines

HEK293T (ATCC, CRL-3216), HCT116 (ATCC, CCL-247), RPE-1 (ATCC, CRL-4000), and Huh-7 (Cellosaurus, CVCL_0336) cells were cultured and passaged in Dulbecco's Modified Eagle Medium (Gibco) supplemented with 10% fetal bovine serum (Gibco) and 1% Penicillin Streptomycin. JeKo-1 (ATCC, CRL-3006) and Jurkat (ATCC, TIB-152) cells were maintained in RPMI-1640 medium (Gibco) supplemented with 10% FBS and 1% Penicillin Streptomycin. All cell lines were incubated at 37 °C in a humidified atmosphere containing 5% CO_2.

Transfection and genomic DNA extraction

For transfection, cells were seeded onto 24-well plates (JETBIOFIL) and transfected at approximately 70% confluence. In fluorescent reporter assays, HEK293T cells were transfected with 1,200 ng of editor plasmid and 500 ng of split-GFP plasmids using EZtrans Cell Transfection Reagent (LIFE iLAB BIO), following the manufacturer's protocol. Flow cytometry was performed 48 hours post-transfection to assess GFP fluorescence. For endogenous site editing, in the IS621 system, HEK293T and HCT116 cells were transfected with 1200 ng of an editor-bRNA all-in-one plasmid and 500 ng of a circular donor plasmid. For experiments using tbRNA or tebRNA, an additional 300 ng of a Csy4 expression plasmid was co-transfected. In the ISCro4 system, HEK293T cells were transfected with a total of 1200 ng of editor

plasmids (two editor-split bRNA all-in-one plasmids, 600 ng each) and 500 ng of a circular donor plasmid. In the evoCAST system, HEK293T cells were transfected with plasmids encoding 100 ng Cas6, 100 ng Cas7, 100 ng Cas8, 200 ng TnsAB, 50 ng TnsC, and 100 ng TniQ, along with 600 ng of a crRNA-donor all-in-one plasmid and 50 ng of a GFP expression plasmid. All transfections were performed using EZtrans Cell Transfection Reagent (LIFE iLAB BIO) according to the manufacturer's instructions. Huh-7 and RPE-1 cells were transfected with the same plasmid amounts using Lipofectamine™ 3000 Transfection Reagent (Invitrogen), following the manufacturer's protocol. Jurkat cells were transfected with the same plasmid combination using the ProteanFect™ Transfection Kit (Nanoportal Biotech), according to the manufacturer's instructions. On day 4 post-transfection, cells were subjected to fluorescence-activated cell sorting (FACS), and the top 20% mCherry-positive population was collected. Genomic DNA was extracted using the Genomic DNA Extraction Kit (BayBio) following the manufacturer's instructions.

Cell-cycle arrest

REP-1 cells were seeded in 24-well plates and cultured until reaching approximately 50% confluence. Cells were then treated with aphidicolin at a final concentration of 2 μg/mL. After 8 h of treatment, cells were subjected to EdU incorporation assays (Meilunbio) for cell cycle analysis or used for plasmid transfection as described above. Three days after transfection, the drug was removed and replaced with normal culture medium, and positive cells were analyzed by flow cytometry on day 4 to determine editing efficiency.

Targeted deep sequencing

Genomic DNA was extracted as previously described. Target loci were amplified using Phanta Max Super-Fidelity DNA Polymerase (Vazyme) with site-specific primers listed in Supplementary Data 5. The resulting PCR amplicons, each tagged with unique barcodes, were pooled and subjected to paired-end deep sequencing (2 x PE150) on the Illumina NextSeq 500 platform (~20 M reads per sample) at Novogene Co., Ltd. (Beijing, China). Insertion efficiencies and indel frequencies at each target site were analyzed using the CRISPResso2 software in batch mode.

RT-qPCR analysis

Standard plasmids were synthesized by Azenta Life Sciences and used to generate a standard curve. Genomic DNA extracted as described above served as the template for absolute quantification of insertion efficiency by RT-qPCR. The reactions were prepared using ArtiCanCEO SYBR qPCR Mix with a total volume of 20 μL, comprising 10 μL SYBR premix, 0.8 μL of each forward and reverse primer, 7.4 μL ddH_2O, and 1 μL DNA template. RT-qPCR was performed on a QuantStudio™ 1 Plus Real-Time PCR system using the following thermal cycling conditions: initial denaturation at 95 °C for 5 min, followed by 40 cycles of denaturation at 95 °C for 15 s, annealing and extension at 60 °C for 20 s, and elongation at 72 °C for 20 s. A melt curve analysis was conducted at the end of the amplification to verify specificity. To establish the standard curve, the plasmid standards were serially diluted in 10-fold gradients. Absolute copy numbers of the target gene in each sample were calculated based on the standard curve. The RPPH1 gene was used as an internal reference for normalization. Primer sequences are listed in Supplementary Data 5.

ddPCR analysis

Digital droplet PCR (ddPCR) was performed to quantify the target gene in the above samples. Each reaction (20 μL total volume) was assembled using 2 × T5 Fast qPCR Mix (Probe) and TaqMan probes labeled with either FAM or VIC fluorophores. After mixing all components, droplet generation and PCR amplification were carried out on the DQ24 digital PCR system. The thermal cycling conditions were as follows: initial denaturation at 95 °C for 2 min, followed by 40 cycles of 95 °C for 10 s and 60 °C for 60 seconds. Upon completion, the system read the fluorescence signal of each droplet to determine the number of positive droplets, from which the absolute copy number of both the target and reference genes per ng of DNA was calculated. Copy number variation (CNV) was then determined accordingly. The RPPH1 gene was used as an internal reference for normalization. Primer sequences are listed in Supplementary Data 5.

Pull-down assay

For DNA pull-down assays, 44-bp single-stranded DNA (ssDNA) oligonucleotides containing either the target sequence or the donor sequence (4 oligos) were synthesized separately. The up strand of tDNA (TACGATGCTAGTACGGGGACCCCTAGCTGGCTTACGAGTCCGTATTAA) and dDNA (TACGATGCTAGTACGACAGTATCTTGTATGCTTACGAGTCCGTATTAA) were subsequently labeled at the 3′ end with biotin-dUTP with a commercial kit (Beyotime) according to the manufacturer's instructions. Following biotin labeling, complementary strands were annealed to generate double-stranded tDNA and dDNA. HEK293T cells transfected with IS621, IS621-H138R, or enIS621 (together with bRNA) were lysed using lysis buffer (Coolaber) supplemented with DTT (Aladdin), RNase inhibitor (Beyotime), and PMSF (Beyotime). Cell lysates were clarified by centrifugation, and protein concentrations were determined using a BCA protein assay kit (Beyotime). Equal amounts of biotinylated tDNA or dDNA were incubated with streptavidin magnetic beads, followed by the addition of equal amounts (50 μg) of protein lysate for pull-down assays. All pull-down procedures were performed according to the manufacturer's instructions (Beyotime). For biotinylated RNA pull-down assays, the 177-nt bRNA cassette was amplified by PCR with a T7 promoter sequence added to the 5′ end. The purified PCR product was subjected to in vitro transcription to prepare bRNA, which was subsequently biotinylated (RNA-labeling kit, Beyotime) according to the manufacturer's instructions. HEK293T cells transfected with IS621, IS621-A119R, or enIS621 (no bRNA) were lysed under the same conditions as described for the DNA pull-down assays to obtain lysates expressing, followed by protein quantification using a BCA assay (Beyotime). Equal amounts of biotinylated bRNA (3 μg) were incubated with streptavidin magnetic beads, after which equal amounts (50 μg) of protein lysate were added for pull-down assays. All steps were performed according to the manufacturer's instructions (Beyotime).

Western blot

Protein samples obtained from pull-down assays or cycloheximide assays were mixed with 4× SDS loading buffer (Beyotime) and denatured at 95 °C for 10 min. Proteins were separated by SDS-PAGE and subsequently transferred onto PVDF membranes using a wet transfer system. After transfer, membranes were blocked in TBST buffer containing BSA for 1 h at room temperature. Membranes were then incubated with the MYC tag Polyclonal antibody (Proteintech, 16286-1-AP; diluted 1:2000) or GAPDH Polyclonal antibody (Proteintech, 10494-1-AP; diluted 1:5000) overnight at 4 °C. Following three washes with TBST, membranes were incubated with HRP-conjugated secondary antibodies (Proteintech, SA00001-2; diluted 1:5000) for 1 hour at room temperature. After additional washes with TBST, protein signals were detected using enhanced chemiluminescence reagents (Millipore) and imaged using a chemiluminescence detection system. Band intensities were quantified using ImageJ software.

Long read sequencing analysis of integration products

To evaluate the structural integrity of exogenous DNA integration products, long-read sequencing was performed at the target locus. Following cell transfection and genome editing as described above, genomic DNA was extracted from edited cells, and long fragments

spanning the integration region were amplified. Two sets of PCR primers (Amplicon 1 and Amplicon 2) were designed to amplify either half of the integrated region together with the flanking genomic sequences. The placements of the donor-specific primers would lead to a significant overlap between the two PCR amplicons. The PCR products were next purified and subjected to PacBio circular consensus sequencing (CCS) (Annoroad, Beijing, China). The resulting CCS reads were first demultiplexed according to the sequencing barcodes and PCR primer sequences, and then aligned to a reference sequence containing the target genomic region and the donor DNA. Sequence alignment was performed using minimap2, followed by sorting and indexing of the alignment files with samtools. Only reads with primary alignment and mapping quality ≥ 20 were retained for downstream analysis. Indels events were extracted from the CIGAR strings of aligned reads, and their positions in the reference sequence were recorded. For each nucleotide position across the integration region, the sequencing coverage, the number of reads containing indels, and the corresponding indel frequency (indel reads divided by total coverage) were calculated. A per-base indel profile across the integration region was then generated to assess the sequence integrity and structural stability of the donor DNA and its junctions with the genomic DNA.

Genome-wide off-target integration analysis

Plasmids expressing enIS621-bRNA, enIS621-ebRNA, enIS621-tbRNA, or enIS621-tebRNA were co-transfected with a 2.3 kb circular donor DNA into cells. Two independent biological replicates were performed for each experimental group. Four days after transfection, fluorescently positive cells were isolated by flow cytometric sorting, and genomic DNA was extracted. To capture donor-genome junction sequences, 100 ng of genomic DNA was randomly fragmented and ligated with sequencing adapters using Tn5 Transposome (Beyotime) according to the manufacturer's instructions. Two rounds of PCR amplification were subsequently performed to enrich fragments containing integration junctions. In the first round of PCR (12 cycles), an outer donor-specific primer was used to amplify DNA fragments containing donor-genome junctions. In the second round of PCR (12 cycles), a nested donor-specific primer containing Illumina adapter sequences was used for further amplification. PCR products were separated by agarose gel electrophoresis, and DNA fragments ranging from 250 to 750 bp were gel-purified and subjected to paired-end Illumina sequencing. Sequencing reads were first subjected to quality control and removal of adapter sequences. Donor sequences were then searched within the reads and the downstream genomic flanking sequences were extracted. Specifically, local alignment was performed using edlib, allowing up to four edit distances, including mismatches or indels, to detect the donor sequence or its reverse complement. For reads containing donor sequences, the downstream genomic region adjacent to the donor was extracted, and only flanking sequences of at least 30 bp were retained for further analysis. The extracted flanking sequences were subsequently aligned to the human reference genome GRCh38 using BWA-MEM. For successfully mapped reads, the 5′ end position of each read was defined as the genomic integration site. To reduce the impact of sequencing errors or minor positional shifts, integration sites located within ±5 bp were clustered and considered as a single integration event. To increase confidence in the identified sites, only integration sites detected in both biological replicates were retained, and a read frequency threshold of 0.015% was applied. The on-target site at chr14:22766901 was excluded from the final analysis. The genomic distribution of integration sites was visualized using Circos plots.

Genome-wide off-target cleavage analysis

Electroporation was performed using the Lonza SE Cell Line 4D-Nucleofector™ X Kit according to the manufacturer's protocol on the Lonza 4D-Nucleofector™ system. Four days after electroporation, fluorescently positive cells were isolated by flow cytometric sorting, and genomic DNA was extracted using a genomic DNA extraction kit (BayBio). Library construction and sequencing were carried out by Azenta Life Sciences. Briefly, genomic DNA was randomly fragmented and subjected to end repair and adaptor ligation. PCR amplification was performed using primers specific to the oligonucleotide tag (ODN), and the resulting libraries were sequenced on the Illumina platform. Sequencing reads were quality-filtered and adaptor-trimmed using cutadapt, and PCR duplicates were removed based on UMI tags. Clean reads were aligned to the reference genome using bwa, and off-target sites were identified using GUIDE-seq-specific scripts. Genomic annotation of cleavage sites was performed with ANNOVAR, yielding the genome-wide distribution and annotation of DSBs across chromosomes.

Off-target analysis

HEK293T cells were transfected as described above. Potential off-target sites identified by GUIDE-seq were amplified using site-specific primers and Phanta Max Super-Fidelity DNA Polymerase (Vazyme). PCR products containing unique barcodes were pooled and subjected to deep sequencing on the Illumina NextSeq 500 platform (2 × PE150; ~20 M reads per sample) at Novogene Co., Ltd. (Beijing, China). For each candidate site, indel frequencies were analyzed using the CRISPResso2 pipeline in batch mode. Primer sequences used for off-target site amplification are listed in Supplementary Data 6.

Flow cytometric analysis of CD19-CAR expression

To assess the surface expression of CD19-CAR on Jurkat cells, an indirect flow cytometry approach was employed using biotinylated recombinant human CD19 protein and fluorophore-conjugated streptavidin. Briefly, Jurkat cells transfected with the recombination mix (corresponding to targeted CAR insertion) were cultured under puromycin selection for three weeks. The surviving cells were harvested, washed, and incubated with biotinylated recombinant human CD19 protein (Yeasen) according to the manufacturer's instructions, followed by washing to remove unbound protein. Cells were then incubated with YSFluor™ 488-conjugated streptavidin (Yeasen) following the manufacturer's protocol and washed again. Finally, cells were resuspended in PBS and analyzed using a flow cytometer equipped with a 488 nm laser. Non-transduced Jurkat cells were used as a negative control to determine the percentage of CAR-positive cells.

ELISAs

To assess IL-2 secretion, wild-type Jurkat cells or CD19 CAR-Jurkat cells (after puromycin selection for three weeks) were co-cultured with CD19-expressing JeKo-1 target cells in 96-well plates at an effector-to-target (E:T) ratio of 1:1 (1×10^5 cells per well for both cell types). Each well received 200 µL of RPMI-1640 complete medium supplemented with 10% fetal bovine serum and was incubated at 37 °C in a humidified 5% CO_2 atmosphere for 16 hours. Following incubation, supernatants were collected and centrifuged at 700 × *g* for 5 minutes to remove cells. The resulting cell-free supernatants were assayed using a high-sensitivity human IL-2 ELISA kit (Beyotime), following the manufacturer's instructions. Absorbance values (OD) were measured at 450 nm using a microplate reader. IL-2 secretion levels in each experimental group were expressed as fold change relative to the control group (Jurkat-WT plus JeKo-1). To evaluate factor IX expression, wild-type Huh-7 cells or *F9*-integrated Huh-7 cells (after puromycin selection for three weeks) were seeded at 1×10^5 cells per well in 96-well plates and cultured under standard conditions (37 °C, 5% CO_2) for 48 hours. At the end of the incubation, culture supernatants were harvested and centrifuged at 700 × *g* for 5 minutes to remove residual cells. The concentration of secreted factor IX protein was

determined using a human coagulation factor IX ELISA kit (Sangon Biotech), following the manufacturer's protocol. Absorbance at 450 nm was measured using a microplate reader. Factor IX expression levels were reported as fold change relative to the wild-type Huh-7 control.

Statistics & Reproducibility

Except for Western blot, modified UDiTaS, GUIDE-seq, and ddPCR data, all individual data points presented represent summary results from three biological replicates. Data analysis and graphing were performed using GraphPad Prism version 10. Data are presented as mean ± S.D., statistical significance between two groups was determined using unpaired, two-tailed Student's t-tests. No statistical method was used to predetermine sample size. No data were excluded from the analyses. The experiments were not randomized. The investigators were not blinded to allocation during experiments or outcome assessment.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Modified UDiTaS data, GUIDE-seq, and deep sequencing data have been deposited in the National Center of Biotechnology Information Sequence Read Archive under accession code PRJNA1307969. Source data are provided with this paper.

References

1. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
2. Gaudelli, N. M. et al. Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).
3. Anzalone, A. V. et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).
4. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
5. Mali, P. et al. RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
6. Heyer, W. D., Ehmsen, K. T. & Liu, J. Regulation of homologous recombination in eukaryotes. *Annu. Rev. Genet.* **44**, 113–139 (2010).
7. Suzuki, K. et al. In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature* **540**, 144–149 (2016).
8. Cox, D. B., Platt, R. J. & Zhang, F. Therapeutic genome editing: prospects and challenges. *Nat. Med.* **21**, 121–131 (2015).
9. Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* **36**, 765–771 (2018).
10. Haapaniemi, E., Botla, S., Persson, J., Schmierer, B. & Taipale, J. CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. *Nat. Med.* **24**, 927–930 (2018).
11. Wang, C. et al. dCas9-based gene editing for cleavage-free genomic knock-in of long sequences. *Nat. Cell Biol.* **24**, 268–278 (2022).
12. Yarnall, M. T. N. et al. Drag-and-drop genome insertion of large sequences without double-strand DNA cleavage using CRISPR-directed integrases. *Nat. Biotechnol.* **41**, 500–512 (2023).
13. Anzalone, A. V. et al. Programmable deletion, replacement, integration and inversion of large DNA sequences with twin prime editing. *Nat. Biotechnol.* **40**, 731–740 (2022).
14. Zhang, R. et al. Amplification editing enables efficient and precise duplication of DNA from short sequence to megabase and chromosomal scale. *Cell* **187**, 3936–3952 e3919 (2024).
15. Lampe, G. D. et al. Targeted DNA integration in human cells without double-strand breaks using CRISPR-associated transposases. *Nat. Biotechnol.* **42**, 87–98 (2024).
16. Witte, I. P. et al. Programmable gene insertion in human cells with a laboratory-evolved CRISPR-associated transposase. *Science* **388**, eadt5199 (2025).
17. Querques, I. et al. A highly soluble Sleeping Beauty transposase improves control of gene insertion. *Nat. Biotechnol.* **37**, 1502–1512 (2019).
18. Yusa, K., Zhou, L., Li, M. A., Bradley, A. & Craig, N. L. A hyperactive piggyBac transposase for mammalian applications. *Proc. Natl. Acad. Sci. USA.* **108**, 1531–1536 (2011).
19. Durrant, M. G. et al. Bridge RNAs direct programmable recombination of target and donor DNA. *Nature* **630**, 984–993 (2024).
20. Hiraizumi, M. et al. Structural mechanism of bridge RNA-guided recombination. *Nature* **630**, 994–1002 (2024).
21. Siddiquee, R., Pong, C. H., Hall, R. M. & Ataide, S. F. A programmable seekRNA guides target selection by IS1111 and IS110 type insertion sequences. *Nat. Commun.* **15**, 5235 (2024).
22. Tao, J., Wang, Q., Mendez-Dorantes, C., Burns, K. H. & Chiarle, R. Frequency and mechanisms of LINE-1 retrotransposon insertions at CRISPR/Cas9 sites. *Nat. Commun.* **13**, 3685 (2022).
23. Chen, Y. et al. All-RNA-mediated targeted gene integration in mammalian cells with rationally engineered R2 retrotransposons. *Cell* **187**, 4674–4689 e4618 (2024).
24. Liu, Y. et al. Enhancing prime editing by Csy4-mediated processing of pegRNA. *Cell Res.* **31**, 1134–1136 (2021).
25. Andreu, A. L. et al. A nonsense mutation (G15059A) in the cytochrome b gene in a patient with exercise intolerance and myoglobinuria. *Ann. Neurol.* **45**, 127–130 (1999).
26. Suma, A., Coronel, L., Bussi, G. & Micheletti, C. Directional translocation resistance of Zika xrRNA. *Nat. Commun.* **11**, 3749 (2020).
27. Zhao, M. & Woodside, M. T. Mechanical strength of RNA knot in Zika virus protects against cellular defenses. *Nat. Chem. Biol.* **17**, 975–981 (2021).
28. Liu, G., Yin, K., Zhang, Q., Gao, C. & Qiu, J. L. Modulating chromatin accessibility by transactivation and targeting proximal dsgRNAs enhances Cas9 editing efficiency in vivo. *Genome Biol.* **20**, 145 (2019).
29. Rega, C. et al. High resolution profiling of cell cycle-dependent protein and phosphorylation abundance changes in non-transformed cells. *Nat. Commun.* **16**, 2579 (2025).
30. Pelea, O. et al. Programmable genome editing in human cells using RNA-guided bridge recombinases. *Science* **391**, eadz1884 (2026).
31. Perry, N. T. et al. Megabase-scale human genome rearrangement with programmable bridge recombinases. *Science* **391**, eadz0276 (2026).
32. Giannoukos, G. et al. UDiTaS, a genome editing detection method for indels and genome rearrangements. *BMC Genomics* **19**, 212 (2018).
33. Durrant, M. G. et al. Systematic discovery of recombinases for efficient integration of large DNA sequences into the human genome. *Nat. Biotechnol.* **41**, 488–499 (2023).
34. Finney, H. M., Lawson, A. D., Bebbington, C. R. & Weir, A. N. Chimeric receptors providing both primary and costimulatory signaling in T cells from a single gene product. *J. Immunol.* **161**, 2791–2797 (1998).
35. Imai, C. et al. Chimeric receptors with 4-1BB signaling capacity provoke potent cytotoxicity against acute lymphoblastic leukemia. *Leukemia* **18**, 676–684 (2004).
36. Maude, S. L. et al. Tisagenlecleucel in children and young adults with B-cell lymphoblastic leukemia. *N. Engl. J. Med.* **378**, 439–448 (2018).

37. Porter, D. L., Levine, B. L., Kalos, M., Bagg, A. & June, C. H. Chimeric antigen receptor-modified T cells in chronic lymphoid leukemia. *N. Engl. J. Med.* **365**, 725–733 (2011).
38. Li, H. S. et al. High-performance multiplex drug-gated CAR circuits. *Cancer Cell* **40**, 1294–1305 e1294 (2022).
39. Zhang, J. et al. Non-viral, specifically targeted CAR-T cells achieve high safety and efficacy in B-NHL. *Nature* **609**, 369–374 (2022).
40. Castaman, G. & Matino, D. Hemophilia A and B: molecular and clinical similarities and differences. *Haematologica* **104**, 1702–1709 (2019).
41. George, L. A. et al. Hemophilia B gene therapy with a high-specific-activity factor IX variant. *N. Engl. J. Med.* **377**, 2215–2227 (2017).
42. Pipe, S. W. et al. Gene Therapy with Etranacogene Dezaparvovec for Hemophilia B. *N. Engl. J. Med.* **388**, 706–718 (2023).
43. Wang X. et al. Long sequence insertion via CRISPR/Cas gene-editing with transposase, recombinase, and integrase. *Curr. Opin. Biomed. Eng.* **28**, 100491 (2023).
44. Lee, J., Kweon, J. & Kim, Y. Emerging trends in prime editing for precision genome editing. *Exp. Mol. Med.* **57**, 1381–1391 (2025).
45. Chen, C. et al. Expanding the frontiers of genome engineering: A comprehensive review of CRISPR-associated transposons. *Biotechnol. Adv.* **78**, 108481 (2025).
46. Sandoval-Villegas N., Nurieva W., Amberger M., Ivics Z. Contemporary transposon tools: a review and guide through mechanisms and applications of Sleeping Beauty, piggyBac and Tol2 for genome engineering. *Int. J. Mol. Sci.* **22**, 5084 (2021).
47. Faure, G. et al. TIGR-Tas: A family of modular RNA-guided DNA-targeting systems in prokaryotes and their viruses. *Science* **388**, eadv9789 (2025).
48. Vaysset, H., Meers, C., Cury, J., Bernheim, A. & Sternberg, S. H. Evolutionary origins of archaeal and eukaryotic RNA-guided RNA modification in bacterial IS110 transposons. *Nat. Microbiol.* **10**, 20–27 (2025).

Acknowledgements

We thank the members of the Wu, Liu, Wang & Huang lab for their contributions in providing experimental materials and engaging in valuable discussions. We also thank the Molecular and Cellular Platform of the School of Life Science and Technology (SLST) at ShanghaiTech University and the Drug Discovery Platform of the Shanghai Institute for Advanced Immunochemical Studies (SIAIS) at ShanghaiTech University for their technical support.

Author contributions

L.W., J.L., X.W., and X.H. designed, conceived, and supervised the work. J.G., Z.S., and W.W. performed the experiments and analyzed the data with the assistance of S.H., Y.Z., and G.L. J.G., L.W., J.L., X.W., and X.H. wrote and revised the manuscript with the input of all authors.

Funding

This work was supported by the National Natural Science Foundation of China (Grant no. 82450113), the National Key R&D Program of China (2024YFA1306400), and the Jiangsu Basic Research Center for Synthetic Biology Grant (BK20233003).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-026-74164-z.

Correspondence and requests for materials should be addressed to Xiaolong Wang, Jianghuai Liu or Lina Wu.

Peer review information *Nature Communications* thanks Bin Liu and the other anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.