

Enhancing homology-directed repair efficiency with HDR-boosting modular ssDNA donor

Received: 2 October 2023

Accepted: 22 July 2024

Published online: 10 August 2024



Ying-Ying Jin^{1,4}, Peng Zhang^{1,4}, Le-Le Liu¹, Xiang Zhao¹, Xiao-Qing Hu¹, Si-Zhe Liu¹, Ze-Kun Li¹, Qian Liu¹, Jian-Qiao Wang¹, De-Long Hao¹, Zhu-Qin Zhang¹, Hou-Zao Chen^{1,2,5}✉ & De-Pei Liu^{1,2,3,5}✉

Despite the potential of small molecules and recombinant proteins to enhance the efficiency of homology-directed repair (HDR), single-stranded DNA (ssDNA) donors, as currently designed and chemically modified, remain sub-optimal for precise gene editing. Here, we screen the biased ssDNA binding sequences of DNA repair-related proteins and engineer RAD51-preferred sequences into HDR-boosting modules for ssDNA donors. Donors with these modules exhibit an augmented affinity for RAD51, thereby enhancing HDR efficiency across various genomic loci and cell types when cooperated with Cas9, nCas9, and Cas12a. By combining with an inhibitor of non-homologous end joining (NHEJ) or the HDRobust strategy, these modular ssDNA donors achieve up to 90.03% (median 74.81%) HDR efficiency. The HDR-boosting modules targeting an endogenous protein enable a chemical modification-free strategy to improve the efficacy of ssDNA donors for precise gene editing.

Precise gene editing is significant for both biological research and clinical gene therapy applications. Gene editing relies on intrinsic DNA repair pathways triggered by DNA damage, such as double-stranded break (DSB), site-specifically induced by programmable endonucleases¹. Endogenous proteins are recruited to these sites, where they repair the DSB via two major pathways: error-prone non-homologous end joining (NHEJ) and precise homology-directed repair (HDR)². When these pathways function ineffectively, other error-prone repair pathways, microhomology-mediated end joining (MMEJ) and single-stranded annealing (SSA), evolved presumably to mediate repair^{1,3,4}. NHEJ is initiated by the binding of the Ku70-Ku80 heterodimer to DSB ends and often induces insertions and deletions (indels). In contrast, HDR utilizes proteins like CtIP, the MRN complex (MRE11-RAD50-NBS1), and RAD51 to repair the DSB precisely, using sister chromatids as templates². Therefore, exogenously provided DNA donors containing the intended sequence can also be integrated at the target site via HDR, enabling precise gene editing^{5,6}. Because the programmable endonucleases, Cas9

and Cas12a, have demonstrated high efficiency in inducing DSB, the relatively low efficiency of HDR compared to NHEJ has been a major bottleneck in achieving precise gene editing at desired loci. Recently, inhibition of error-prone repair pathways by dominant-negative 53BP1-fused Cas9⁷, the small molecule M3814^{8,9}, and HDRobust strategy⁹ have been reported to shift the endogenous DNA repair pathway toward HDR, yielding high precise gene editing efficiency. Nonetheless, the involvement of exogenously delivered DNA donors remains a limiting step for optimal HDR efficiency^{5,6}.

Single-stranded DNA (ssDNA) donors generally exhibit higher HDR efficiency and lower cytotoxicity than double-stranded DNA (dsDNA) donors¹⁰. Therefore, substantial effort has been devoted to enhancing the efficiency of ssDNA donor-mediated HDR^{5,6}, also referred to as single-strand template repair (SSTR). To date, most effective and reliable strategies for enhancing the involvement of ssDNA donors are based on tethering ssDNA donors to Cas9 ribonucleoprotein (RNP) complexes, including Cas9-avidin biotin ssDNA

¹State Key Laboratory of Common Mechanism Research for Major Diseases, Department of Biochemistry and Molecular Biology, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100005, China. ²Medical Epigenetics Research Center, Chinese Academy of Medical Sciences, Beijing 100005, China. ³Haihe Laboratory of Cell Ecosystem, Tianjin 300301, China. ⁴These authors contributed equally: Ying-Ying Jin, Peng Zhang. ⁵These authors jointly supervised this work: De-Pei Liu, Hou-Zao Chen. ✉e-mail: chenhouzao@ibms.cams.cn; liudp@pumc.edu.cn

system¹¹, RNP system¹², Cas9-PCV system¹³, Cas9-AeF DBCO-adaptor ssODN system¹⁴, SImplex system¹⁵, and the gDonor system¹⁶. This indicates that the proximity of the ssDNA donor to the target sites is a key determinant of HDR efficiency. However, these strategies rely on chemical modifications of the Cas9 protein and/or the ssDNA donor, leading to several drawbacks, including unsuitability for mRNA-lipid nanoparticle (LNP) and viral delivery systems, incompatibility with other programmable endonucleases, the potential to reduce Cas9 activity, the inclusion of additional recombinant proteins, gene therapy immunogenicity, and increased costs of producing gene-editing tool. Thus, it is necessary to develop a chemical modification-free approach that can recruit ssDNA donors to the target site and enhance HDR efficiency.

Beyond carrying genetic information, ssDNA can serve as pathogen-associated molecular patterns (PAMPs) to trigger innate immune responses¹⁷, antisense oligonucleotides (ASOs) to regulate gene expression¹⁸, aptamers to modulate protein functions¹⁹, and more. Since these functions rely on the interaction between short, specific ssDNA sequences and endogenous proteins, we hypothesize that incorporating certain functional sequence modules into ssDNA donors could enhance their capabilities by facilitating the recruitment of ssDNA donors to DSB sites through targeted interactions with endogenous proteins (Supplementary Fig. 1).

Here, we explore the single-stranded oligodeoxynucleotides (referred to hereafter as ODNs) binding preference sequences of DSB repair-related proteins, which are recruited to the DSB sites for DNA repair. Based on the RAD51-preferred sequences, we develop HDR-boosting modules and incorporate them into ssDNA donors. The inclusion of these modules in the ssDNA donors increases the efficiency of precise gene editing induced by Cas9, nCas9, and Cas12a endonucleases. Notably, the combination of these modular ssDNA donors with the small molecule M3814 or HDRobust strategy leads to HDR efficiencies at endogenous sites ranging from 66.62% to 90.03%. By targeting endogenous DSB repair-related protein through its preferred binding sequences, our chemical modification-free approach represents a simple and potentially safer method to improve the efficiency of ssDNA donors for precise gene editing compared to other chemical modification-based ssDNA donor tethering strategies.

Results

DSB repair-related proteins bind ODNs in a sequence-biased manner

To elucidate the ODN binding features of DSB repair-related proteins and explore their potential binding preference sequences for the development of HDR-boosting sequence modules, we prioritized proteins that participate in the early stages of the HDR pathway, such as CtIP, RAD50, and RAD51, and selected Ku80 as a representative protein involved in the NHEJ pathway^{20,21}. To assess the ODN-binding abilities of these proteins, we used a biotinylated ODN with 24 randomly assembled nucleotides (nt) and performed a luminous ODN immunoprecipitation (ODIP) assay in HEK 293T cells¹⁷. The antibodies specifically immunoprecipitated their target proteins (Supplementary Fig. 2a), and most proteins showed detectable ODN-binding activities (Supplementary Fig. 2b). To determine whether these proteins bind ODNs in a sequence-biased manner and obtain their binding preference sequences, we performed ODIP-Seq in HEK 293T cells using a synthetic ODN pool (Fig. 1a). The ODN pool was prepared by equally mixing 200 ODNs generated based on the ClinVar database (hereafter termed SSO1 to SSO200)¹⁷, and thus preclude the effect of ODN concentration on the binding activity of candidate proteins. Among the four selected proteins, RAD51 and Ku80 exhibited the highest sequence-biased binding activity (Fig. 1b, c). Moreover, the top-ranked ODNs bound to RAD51 and Ku80 protein were characterized through the WebLogo3 website (Fig. 1d, e).

We further validated the binding activity between preferential ODNs and target proteins using two complementary assays: biotin-ODN pulldown and ODIP. Consistent with the ODIP-Seq results (Fig. 1b), SSO14 and SSO9 were verified as RAD51-preferred ODNs (Fig. 1f, g and Supplementary Fig. 3a), whereas SSO64 and SSO17 demonstrated a greater affinity for Ku80 (Fig. 1h, i). Moreover, the enrichment of these ODNs was not attributed to non-specific IgG binding (Fig. 1g, i) or ODN stability effects in cell lysates during ODIP assay (Supplementary Fig. 3b). In agreement with the sequencing data (Fig. 1d), the “TCCCC” motif in SSO9 and SSO14 was necessary for enhancing RAD51 binding (Supplementary Fig. 3c, d), further supporting the hypothesis that these ODNs harbor RAD51-preferred binding sequences.

To test whether SSO9 and SSO14 motifs can enhance the binding between the ssDNA donor and RAD51, we incorporated these motifs into ssDNA donor and performed biotin-pulldown assays. These motifs did not alter the overall ssDNA-binding protein profile but did enhance the binding between the donor and RAD51 (Supplementary Fig. 3e, f). However, mass spectrometry (MS) did not detect RAD51 from these biotin-pulldown samples (Supplementary Fig. 3g), likely due to the limited sequence throughput of MS for complex total protein samples²².

The optimal interface in ssDNA donors for additional DNA sequence module

To test the feasibility of functional sequence modules for ssDNA donors, our next objective was to assess whether the 5′ or 3′ end of an ssDNA donor harbors tolerance sequences that could serve as a module-installing interface without compromising the ssDNA donor’s ability to serve as an effective DNA repair template. We constructed a single-copy, genomically integrated blue fluorescent protein (BFP) reporter cell model to assess the potency of ssDNA donor for mediating HDR, which can implement BFP to green fluorescence protein (GFP) conversion when the 66th amino acid “His” in BFP is replaced by amino acid “Tyr”²³. Conversion of the fluorescent protein required HDR-mediated substitution of the codon, thereby coupling fluorescence to HDR efficiency (Supplementary Fig. 4a). To eliminate clone-to-clone variation caused by the random integration sites of *BFP*, we equally combined four identified clones capable of accurately measuring HDR and indels frequency (Supplementary Fig. 4b–d). Using this model, we tested the HDR efficiencies mediated by ssDNA donors with a series of mutations at each end, and found that the *GFP* ssDNA donor largely maintained its ability to convert *BFP* to *GFP* despite different mutation lengths at the 5′ end, whereas it was more sensitive to the 3′ end mutations (Supplementary Fig. 5a). Consistently, even a single mutant base at 3′ end of the ssDNA donor reduced HDR efficiency (Supplementary Fig. 5b). These findings indicate that the 5′ end is more suitable as an interface to install a functional sequence module.

The RAD51-preferred ODNs enhanced HDR efficiency

As RAD51 is recruited to target sites upon DSB damage^{24,25}, we envisioned that the addition of RAD51-preferred sequences into ssDNA donors might promote their recruitment and improve the HDR efficiency (Fig. 2a). To test this, we incorporated RAD51-preferred SSO9 and SSO14 sequence modules, as well as Ku80-preferred SSO17 and SSO64, into the 5′ ends of the *GFP* donors. Only SSO9 and SSO14 increased HDR efficiency in HEK 293T-BFP reporter cells, while SSO17 and SSO64 did not exhibit enhancement (Fig. 2b and Supplementary Fig. 6a, b), consistent with the respective roles of RAD51 and Ku80 in the HDR and NHEJ pathways^{21,26}. The HDR enhancement effects of SSO9 and SSO14 modules were further validated in K562-BFP reporter cells (Fig. 2c and Supplementary Fig. 6c). Moreover, the enhancement effects were observed at both low and high donor concentrations (Fig. 2d). Hence, we defined SSO9 and SSO14 as HDR-boosting modules for ssDNA donor.

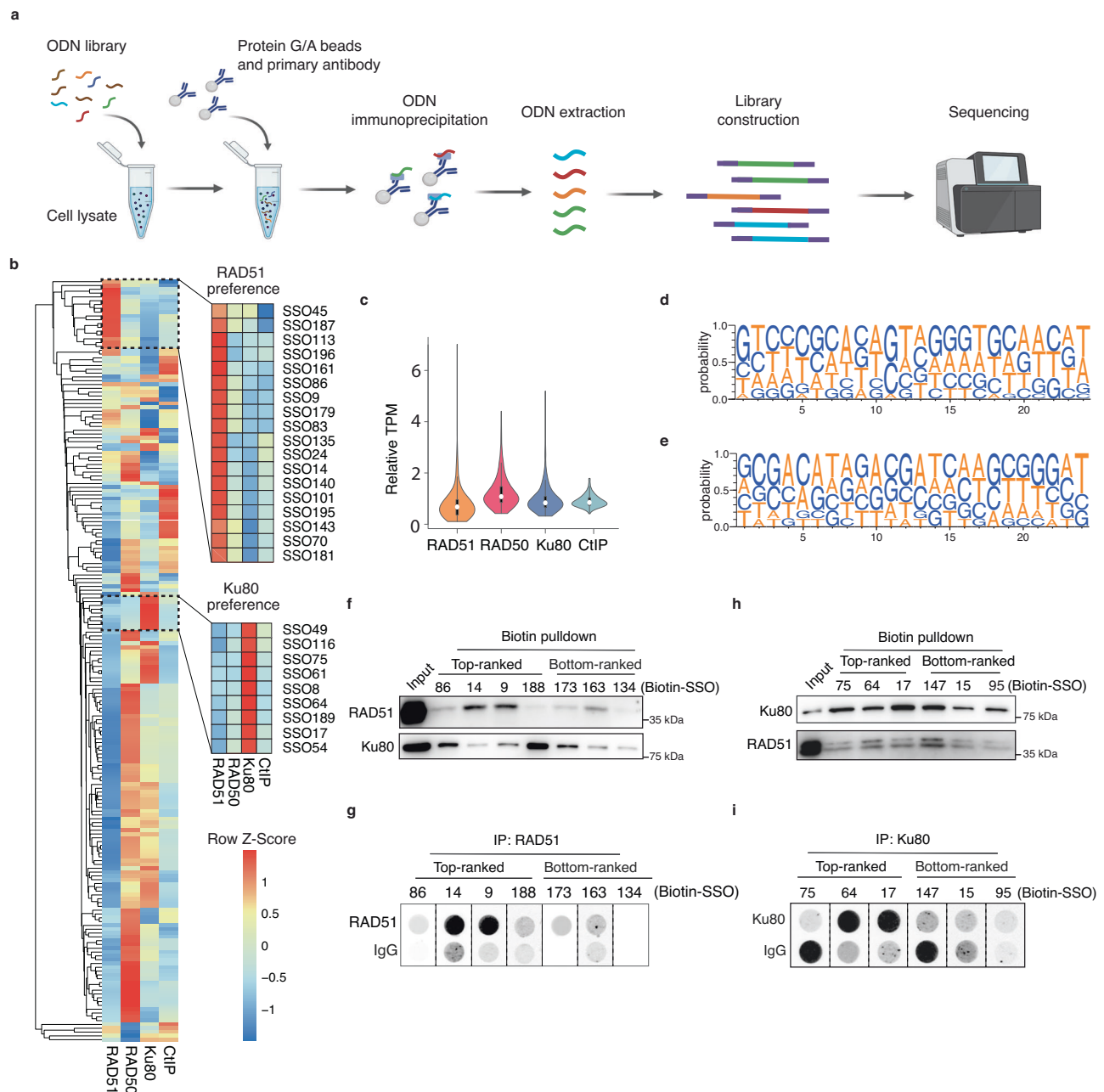


Fig. 1 | DSB repair-related proteins bind ODNs in a sequence-biased manner. **a** Schematic overview of ODIP-Seq for capturing preference binding sequences of the protein of interest. The ODN pool was incubated with the cell lysate for 12 h, and the ODN-protein complexes were immunoprecipitated using the corresponding primary antibody. The ODNs were then recovered from the beads using the proteinase K-phenol-chloroform method, and a library was prepared for next-generation sequencing. ODIP, oligodeoxynucleotide immunoprecipitation. **b** Heatmap of relative TPM of precipitated ODNs with RAD51, RAD50, Ku80 and CtIP. Relative TPM for each ODN was calculated using the mean of two independent replicates. **c** Violin plots showing the overall relative TPM distribution of precipitated ODNs with the indicated DNA repair proteins. The white dot represents the median. The box spans from the 25th percentile (first quartile) to the 75th

percentile (third quartile). The whiskers extend to the smallest and largest data points within 1.5 times the interquartile range (IQR). **d, e** Characterization of top-ranked ODNs bound to the RAD51 (**d**) or Ku80 (**e**) protein in Fig. 1b using the WebLogo 3 website. **f, g** Evaluation of the binding activity of RAD51 protein with top- and bottom-ranked SSO by biotin-ODN pulldown assay (**f**) and ODIP assay (**g**). The digits denote the number of SSOs. **h, i** Evaluation of the binding activity of Ku80 protein with top- and bottom-ranked SSO using biotin-ODN pulldown assay (**h**) and ODIP assay (**i**). The digits denote the number of SSOs. Data are representative of 3 independent experiments (**f–i**). The sequences of all SSOs used are shown in Supplementary Data 3. Source data are provided as a Source Data file. Figure **a** Created with BioRender.com released under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International license.

In line with the previous results (Supplementary Fig. 5), the HDR-boosting modules were tolerated when incorporated at the 5' end but not the 3' end of the ssDNA donor (Fig. 2e and Supplementary Fig. 6d). To ensure that the HDR improvement attributed to the RAD51-preferred sequences rather than the homology arm elongation, we designed a donor with 5' elongated homology arm. Although

this elongated ssDNA donor exhibited a certain increase in HDR efficiency compared to the control donor, its potency was less than the modular donors (Supplementary Fig. 6e). Furthermore, no additional improvement in HDR efficiency was achieved when incorporating two tandem modules into the ssDNA donor (Supplementary Fig. 6f).

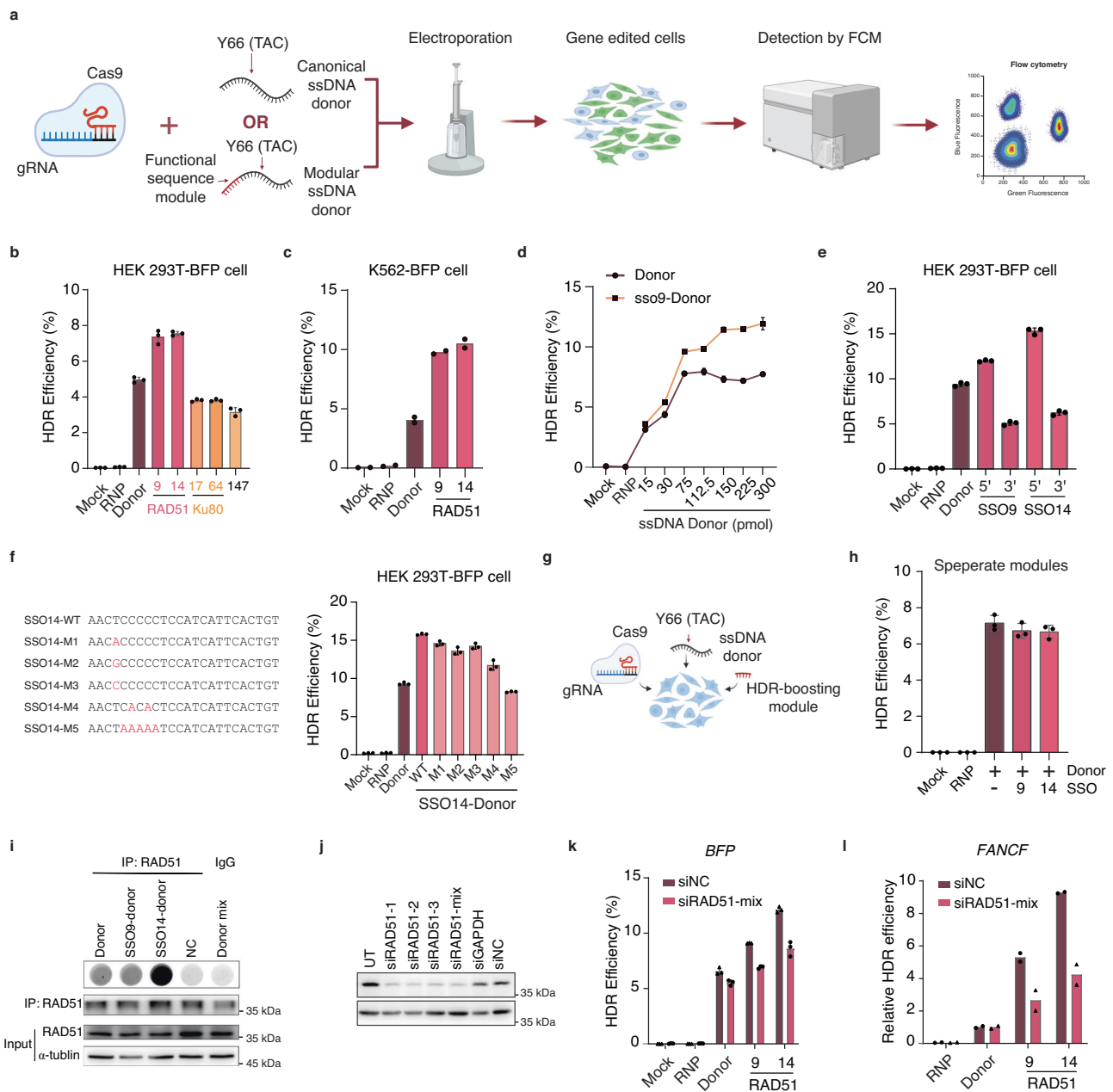


Fig. 2 | Addition of RAD51-preferred ODNs increases ssDNA donor-mediated HDR efficiency at the BFP locus. **a** Schematic representation for gene editing in the BFP reporter cell lines. Single-copy-integrated *BFP* cell lines were electroporated with Cas9 RNP together with modular or canonical ssDNA donors. Then fluorescence conversion percentage was measured using flow cytometry. **b, c** HDR efficiencies of ssDNA donors connected with RAD51- or Ku80-preferred ODNs in HEK 293T-BFP cells (**b**) and K562-BFP cells (**c**). **d** HDR efficiencies of ssDNA donors at different concentration gradients in HEK 293T-BFP cells. **e** The effect of the incorporation position on the HDR improvement of HDR-boosting modules. **f** HDR efficiencies of ssDNA donors incorporated with the indicated mutant SSO14 modules. **g** Schematic representation for gene editing with ssDNA donors and independent HDR-boosting modules. **h** HDR efficiencies of ssDNA donors in HEK 293T-BFP cells electroporated with CRISPR-Cas9 RNP, canonical ssDNA donor, and independent HDR-boosting module. **i** Evaluation of the binding activity of RAD51 with ssDNA donors using the ODIP assay. Donor mix, an equal parts mix of the three

ssDNA donors mentioned in the image. **j** Western blot analysis of the knockdown efficiencies of three siRNAs targeting RAD51 in HEK 293T cells 48 h post-transfection. siRAD51-mix, an equal parts mix of the three siRNAs of RAD51 mentioned in the image. **k, l** Effects of RAD51 knockdown on HDR efficiencies of ssDNA donors at the *BFP* site in HEK 293T-BFP cells (**k**) and endogenous *FANCF* site in HEK 293T cells (**l**). For all HDR efficiency-assessing experiments unless otherwise specified, 18 pmol Cas9 nuclease, 22 pmol gRNA and 6 pmol ssDNA donors corresponded to 2×10^5 cells. HDR efficiency was measured three days after electroporation. Data are representative of 3 independent experiments (**i, j**). Values and error bars reflect mean $\pm$ SD of $n = 3$ (**b, d-f, h, k**) independent electroporation replicates. Values reflect $n = 2$ (**c, l**) independent electroporation replicates. The sequences of all gRNAs, ssDNA donors and siRNAs used are shown in Supplementary Data 1, 2, and 6. Source data are provided as a Source Data file. Figure (**a, g**) Created with BioRender.com released under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International license.

Consistent with the biotin pulldown results (Supplementary Fig. 3c, d), the “TCCCC” motif was crucial for the HDR-boosting module to improve HDR efficiency, as the mutations in this motif impaired the improvement effects (Fig. 2f and Supplementary Fig. 6g).

This result further affirmed that the functional sequences, rather than elongated length, led to the enhancement in HDR efficiency. Notably, when the HDR-boosting modules were co-transfected without being incorporated into the ssDNA donors, they completely lost their HDR-

boosting effects (Fig. 2g, h), indicating that incorporation was necessary for these modules to enhance the ssDNA donor potency. To confirm that the improved HDR efficiency provided by the HDR-boosting modules was indeed due to the enhanced interaction with the RAD51 protein, we employed an ODIP assay and found that the HDR-boosting modules, especially SSO14, promoted the binding of the ssDNA donor to RAD51 (Fig. 2i). Consistent with this observation, RAD51 knockdown partially abrogated the boosting effects of SSO9 and SSO14 (Fig. 2j–l).

HDR-boosting modules function in multiple genomic loci and human cell types

Having validated the HDR-boosting modules as an effective approach to improve HDR frequency in the BFP reporter model, we next investigated their efficacy for interrogating endogenous genes. We installed the modules at the 5' ends of ssDNA donors targeting six endogenous gene loci (*EMX1*, *DNMT1*, *CXCR4*, *RUNX1*, *RNF2*, and *FANCF*)²⁷. The donors were designed to rewrite six base pairs around DSB sites to a Hind III restriction site “AAGCTT” (Supplementary Fig. 7a), we amplified genomic regions outside the homology arms of the ssDNA donors and evaluated HDR frequencies via Hind III cleavage and next-generation sequencing (NGS) complementarily. Both approaches showed improved HDR efficiency with SSO9-connected ssDNA donors versus canonical donors at all examined loci (Supplementary Fig. 7b, c). Compared to Hind III cleavage assays, the NGS approach could more precisely and sensitively provide information on all the specific editing events, such as HDR and indels. Therefore, we analyzed the gene editing efficiency using NGS in the subsequent experiments. Aligning with the results observed in the BFP model (Supplementary Fig. 6e), the improved HDR efficiency conferred by HDR-boosting modules attributed to the functional sequences themselves rather than the elongated 5' homology arm of ssDNA donor (Supplementary Fig. 7d).

Since HDR efficiency varies substantially between mammalian cell types, we tested the potency of the HDR-boosting modules across cell types, including HEK 293T, HeLa, U2OS, and K562 cells with the same base substitution assay. Both HDR-boosting modules (SSO9 and SSO14) exhibited notable abilities at almost all examined genomic loci in these four cell lines, with about 1.7–4.8-fold higher editing (from 4.1–10.5% to 9.4–29.7%) in HEK 293T cells (Fig. 3a and Supplementary Fig. 8a, b), 1.3–15.2-fold higher editing (from 0.2–1.3% to 0.9–9.8%) in HeLa cells (Fig. 3b and Supplementary Fig. 8c, d), 1.1–8.0-fold higher editing (from 1.4–15.5% to 8.2–23.8%) in U2OS cells (Fig. 3c and Supplementary Fig. 8e, f) and 1.8–4.8-fold higher editing (from 4.4–14.5% to 17.9–32.6%) in K562 cells (Fig. 3d and Supplementary Fig. 8g, h), without apparently affecting Cas9 cleavage (Supplementary Fig. 8b, d, f, h). HDR-boosting modules demonstrated even increased activity at sites that were more difficult to edit by the canonical ssDNA donor, such as the *FANCF* gene site (Fig. 3a–d and Supplementary Fig. 8). Notably, SSO14 exhibited a more potent capacity to elevate the HDR efficiency at almost all examined sites compared with SSO9 (Fig. 3a–d and Supplementary Fig. 8), which was consistent with its higher RAD51-binding activity (Fig. 1f, g and Fig. 2i). Together, these results indicate that HDR-boosting modules can increase gene-editing efficiency in various cell types.

In contrast to the commonly employed high dose of ssDNA donor (200–500 pmol per million cells)^{23,28}, we used a relatively low dose (30 pmol per million cells) for most of the previous experiments. To further investigate the possibility of achieving higher HDR efficiency with additional HDR-boosting modular donors, we examined the HDR efficiency of ssDNA donors under a series of concentration gradients at the *FANCF* locus in HeLa cells. Strikingly, the HDR efficiency achieved by 15 pmol modular donors was even higher than that of 150 pmol control donors (Fig. 3e). Within the concentrations of the test, the HDR increase in the modular ssDNA donor was greater than that of the

control donor, lending support to a more active recruitment model for the modular ssDNA donor.

We then sought to interrogate two endogenous gene loci simultaneously, an easy-to-edit locus (*RUNX1*) and a hard-to-edit locus (*EMX1*) along with *FANCF*, respectively. HDR-boosting modules simultaneously improved the HDR efficiency at dual sites (Fig. 3f and Supplementary Fig. 9a). Importantly, maximal achieved HDR efficiency at each site was not compromised compared to single-site editing (Fig. 3b, f).

We next extended our findings to primary cells by inducing gene editing in human peripheral blood (hPB) CD34+ cells, an important cell type for treating blood disorders²⁹. Because editing the *FANCF* target was challenging in human cell lines, we primarily evaluated this site in hPB CD34+ cells. In agreement with the findings in cell lines, SSO9 and SSO14 modules improved the HDR frequency from 2% to 2.8% and 4.9% in hPB CD34+ cells, respectively (Fig. 3g and Supplementary Fig. 9b).

The safety of the HDR-boosting module

Given that the delivered ODNs containing preferred sequences of endogenous proteins may bind these proteins with high affinity, there is a potential risk that their occupancy could impact the activity of these proteins. For instance, ODNs with RAD51-preferred sequences might inhibit the RAD51-mediated HDR pathway, while those with Ku80-preferred sequences could hinder the NHEJ pathway. To evaluate the potential risk posed by these ODNs on DNA repair processes, specifically RAD51-preferred ODNs (SSO9 and SSO14) and Ku80-preferred ODNs (SSO17 and SSO64), we electroporated them into HEK 293T cells alongside Cas9 RNP (Supplementary Fig. 10a). The results revealed that none of these ODNs induced noticeable changes in overall indels efficiency or the detailed repair outcomes (Supplementary Fig. 10b, c). Similarly, they did not exhibit significant influences on HDR efficiency (Fig. 2g, h and Supplementary Fig. 10d, e).

Having evaluated the safety of these free ODNs, we sought to assess the potential risks when HDR-boosting modules are incorporated into ssDNA donors, including genome-wide off-target integration, translocation at DSB sites, and the integration or insertion of HDR-boosting modules with homology arm. To assess the off-target integration rate of our modular ssDNA donors and their potential influences on the translocation at DSB sites, we adopted a Tn5-based high-throughput genome-wide sequencing³⁰ to capture off-target integration and translocation amplicons by specific forward primers and Tn5 adapter primers (Fig. 4a). Consistent with the previous work^{10,31}, we detected a less than 2.42% frequency in off-target integration and translocation at both *FANCF* and *RUNX1* loci in HEK 293T cells (Fig. 4b, c). Importantly, the HDR-boosting modular donors induced neither more unintended disruption at other genomic regions nor translocation at the DSB sites compared to canonical donors (Fig. 4c).

As the functional modules add mismatch sequences to the 5' ends of the homology arms of the donors, these sequences might be unexpectedly incorporated into the targeted genome region with the homology arms³² or lead to direct insertion at the DSB sites. To address these concerns, we profiled the frequency of these unintended editing outcomes at four genome loci where the HDR efficiency was evaluated across four types of cells in Fig. 3. Any single base mutation around the 5' homology arm targeting genomic regions was regarded as an integrity disruption. Control donors almost resulted in undetectable integrity disruption at the 5' homology arm targeting regions, while HDR-boosting modular donors could induce less than a 0.06% frequency in disrupting the 5' homology arm targeting regions (Supplementary Fig. 11a). Compared to the benefits in achieving additional HDR efficiency, these low levels of unintended outcomes are acceptable. Although these HDR-boosting modular donors tended to produce slightly more direct insertion at DSB sites with a frequency of less than 1.03% (Supplementary Fig. 11a), these editing outcomes were

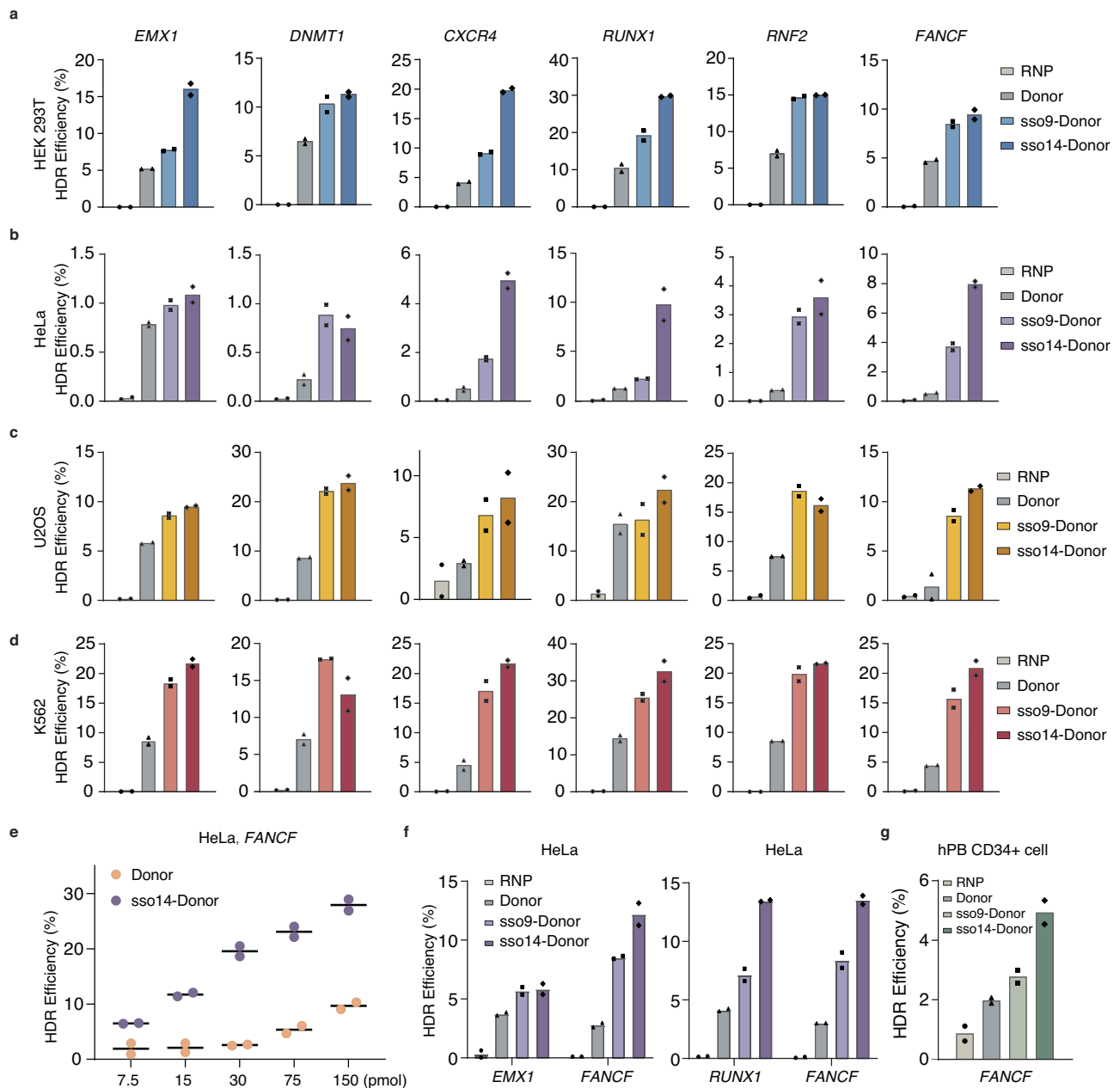


Fig. 3 | HDR-boosting modules achieve efficient precise gene editing at endogenous genomic loci in multiple human cell types. **a–d** HDR efficiencies of HDR-boosting modular ssDNA donors and canonical ssDNA donors at the specified gene loci for *EMX1*, *DNMT1*, *CXCR4*, *RUNX1*, *RNF2*, and *FANCF* in HEK 293T cells (**a**), HeLa cells (**b**), U2OS cells (**c**), and K562 cells (**d**). **e** HDR efficiencies yielded with different concentrations of HDR-boosting modular ssDNA donors and canonical ssDNA donors at the *FANCF* site in HeLa cells. The digits denote the amount of ssDNA donors corresponding to 1×10^6 cells. **f** HDR efficiencies of ssDNA donors interrogating two endogenous gene sites simultaneously in HeLa cells. **g** HDR

efficiencies of HDR-boosting modular ssDNA donors and canonical ssDNA donors at the *FANCF* site in hPB CD34+ cells. For all HDR efficiency-assessing experiments unless otherwise specified, 18 pmol Cas9 nuclease, 22 pmol gRNA and 6 pmol ssDNA donors corresponded to 2×10^5 cells. HDR efficiency was measured by NGS three days after electroporation. HDR efficiencies reflect the sequencing reads that contain the intended edit and do not contain indels among all treated cells (**a–g**). Values reflect $n = 2$ (**a–g**) independent electroporation replicates. The sequences of all gRNAs and ssDNA donors used are shown in Supplementary Data 1 and 2. Source data are provided as a Source Data file.

counted as indels without influencing the intended HDR. Interestingly, we found these unintended editing outcomes were highly correlated to HDR efficiency (Supplementary Fig. 11b), suggesting that both the pronounced HDR and the undesired editing side effects may be attributed to the HDR-boosting module-mediated enhancement of ssDNA donor recruitment to the DSB sites.

Finally, we measured off-target effects at the top four predicted off-target sites of the *FANCF* locus in hPB CD34+ cells. The HDR-boosting modules did not affect the off-target effects of Cas9 RNP, with all ssDNA donors exhibiting less than 1% off-target editing at the

examined sites (Supplementary Fig. 11c). Overall, these results demonstrate that HDR-boosting modules offer a safe and effective strategy for gene editing.

The competence of HDR-boosting modular donors in other types and systems of precise gene editing

After validating that the HDR-boosting modules enhanced base substitution via ssDNA donor-mediated HDR, we assessed if these modules could facilitate inserting a FLAG epitope tag sequence or loxP recognition sequence at the *FANCF* locus in HeLa cells (Fig. 5a). We

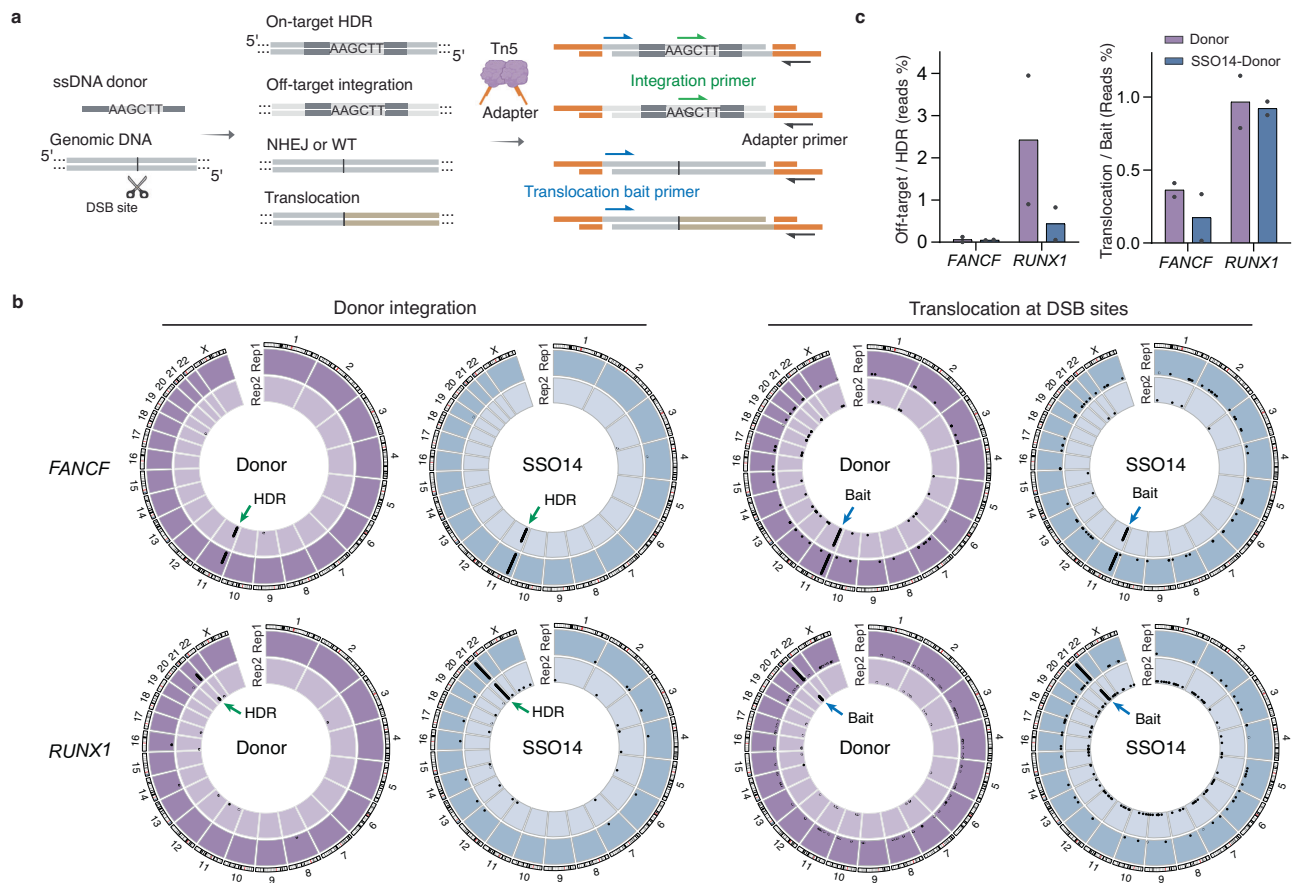


Fig. 4 | Genome-wide profile of donor integration and chromosomal translocation. **a** Brief workflow of Tn5-based high-throughput genome-wide sequencing for integration of ssDNA donor or translocation at the DSB site. Genomic DNA was tagged with Tn5, and genomic regions containing the inserted donor sequences and the translocated regions were amplified using the indicated primer pairs designed for integration and translocation, respectively. After nest-PCR, amplicons were barcoded and then sequenced. **b** Circos plots of genome-wide off-target integrations of donors and translocation junctions in edited cells. Off-target

integrations and translocation junctions were binned to 0.1 Mb regions and plotted on a normalized scale (black bars). Green arrow, ssDNA donor on-target site. Blue arrow, bait primer targeting site. **c** The overall off-target integration and translocation rate were calculated as the percentage of off-target reads in HDR reads and the percentage of translocation reads in non-translocation (bait region) reads, respectively. Values reflect $n = 2$ (**c**) independent electroporation replicates. The sequences of all gRNAs, ssDNA donors and primers used are shown in Supplementary Data 1, 2 and 4. Source data are provided as a Source Data file.

observed an average of 1.3-fold increased efficiency of FLAG tag insertion (from 17.3–24.3% to 20.2–27.4%) and an average of 2.8-fold increased efficiency of loxP insertion (from 4.7–11.5% to 12.3–14.8%) when using the HDR-boosting modular ssDNA donors compared to that using canonical ssDNA donors (Fig. 5b and Supplementary Fig. 9c). The absolute knock-in efficiency reached 27.4%, comparable to the base substitution (Fig. 3e), suggesting that the 5' terminal modules can visibly improve the knock-in efficacy of the ssDNA donor.

An ssDNA donor can also cooperate with paired Cas9 nickases (nCas9) for precise genome editing³³. Therefore, we examined the potency of HDR-boosting modules in this double-nicking system (Fig. 5c). The HDR-boosting module improved HDR efficiency induced by both D10A and H840A nCas9 pairs in HEK 293T-BFP cells, with up to 38.6% HDR efficiency (Fig. 5d). Moreover, PAM-out paired sgRNAs (sgRNA 1 and 3) outperformed PAM-in pairs (sgRNA 1 and 2) when coordinated with different types of nCas9s and donor doses (Fig. 5d). Considering the proteins involved in the following DNA repair pathway of double nicks differ from those involved in the DSB strategy³⁴, further screening modules optimized specifically for the double-nicking system may be worthwhile.

By targeting endogenous RAD51 via HDR-boosting modules, our approach can theoretically be compatible with any programmable nuclease, such as Cas12a, which generates sticky-end DSB that are less prone to NHEJ^{35,36}. Finally, we sought to extend our strategy to this

programmable nucleases-mediated gene editing systems (Fig. 5e). SSO9 and SSO14 modules improved the HDR frequency from 0.2% to 3.9% and 5.1% in HEK 293T cells, respectively (Fig. 5f).

Combination with state-of-the-art strategies for efficient precise gene editing

Donors with certain types of chemical modifications have been reported to increase HDR efficiency by prolonging the half-life of donors and thereby increasing their availability for HDR^{37–39}. Therefore, we introduced biotin, triethyleneglycol (TEG), phosphorothioate (PS) modifications, respectively, to either 5' end or both ends of the ssDNA donors to assess whether these modifications could further enhance HDR efficiency. However, only slight HDR enhancement, from 16.97% to 28.37%, was detected through these strategies (Supplementary Fig. 12a), indicating that the proximity of the ssDNA donor plays a more crucial role in improving HDR efficiency than the half-life time of the ssDNA donor.

Recently, a study reported that the combination of NHEJ inhibitor M3814 and MMEJ inhibition by siRNAs targeting Polθ, a strategy termed HDRobust, achieved notably high HDR efficiency⁹. As the HDRobust strategy relies on the regulation of DSB repair pathways, while our strategy enhances the recruitment of ssDNA donors, we sought to test whether a combination of the HDRobust method and our HDR-boosting modular donors could achieve even higher precise

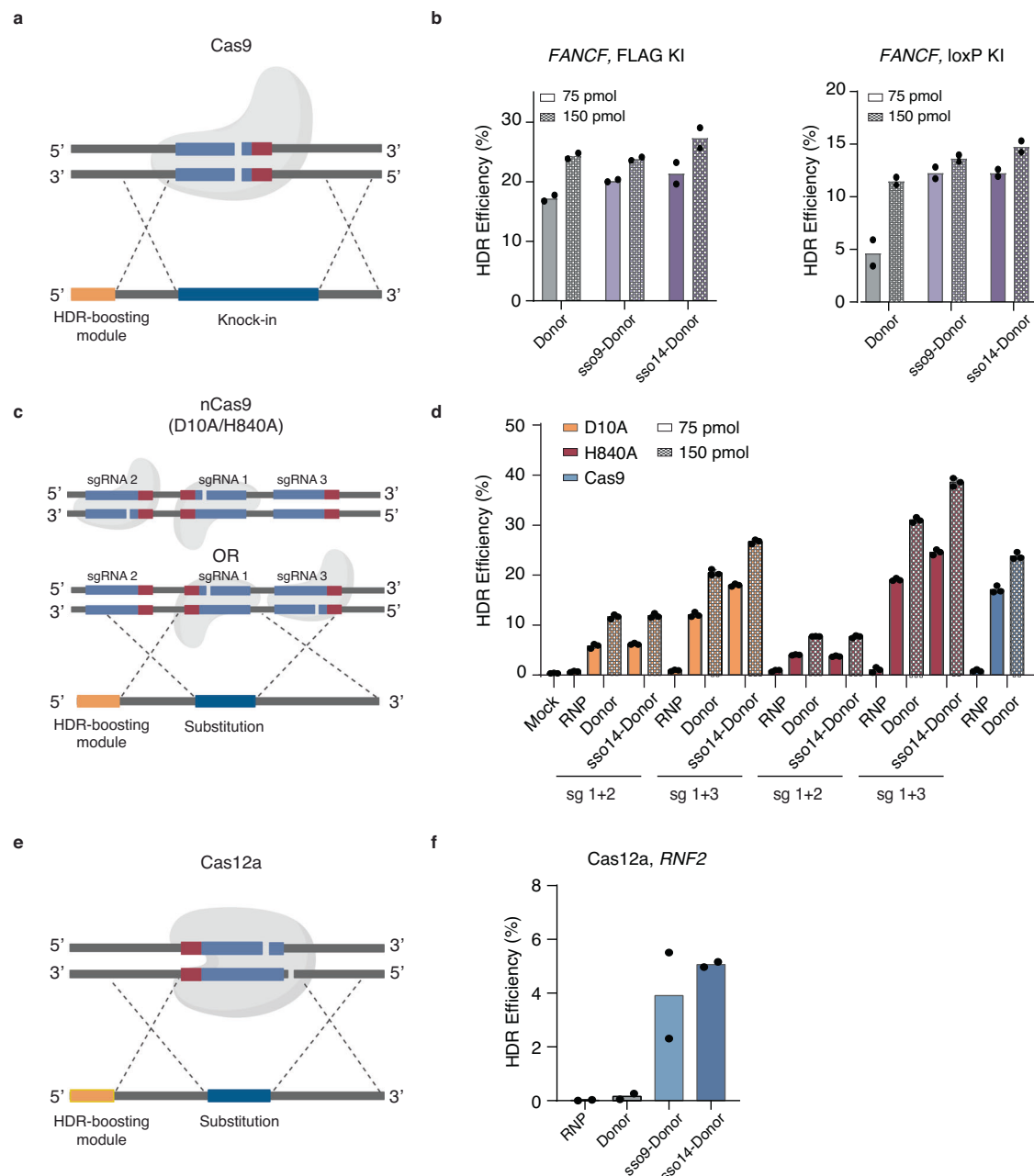


Fig. 5 | HDR-boosting modules function in other types of precise gene editing.

a Schematic representation for DNA knock-in at the endogenous genomic loci cooperating with Cas9. **b** FLAG or loxP sequence insertion efficiency mediated by the indicated ssDNA donors at the *FANCF* gene locus in HeLa cells. 18 pmol Cas9 nuclease, 22 pmol gRNA and ssDNA donors with the indicated amount (per million cells) corresponded to 2×10^5 cells. **c** Schematic illustrating DNA double nicks using a pair of sgRNAs guiding Cas9 nickses (nCas9). The D10A mutation renders Cas9 able to cleave only the strand complementary to the sgRNA; the H840A mutation renders Cas9 able to cleave only the non-complementary strand. A pair of sgRNA-nCas9 complexes can nick both strands simultaneously. **d** HDR efficiencies in paired nCas9-mediated precise editing system using ssDNA donors tethered with HDR-boosting module. In this electroporation, 36 pmol nCas9 nuclease, 22 pmol gRNA, 22 pmol AS-gRNA and the indicated amount of ssDNA donor (per million cells) corresponded to

2×10^5 cells. **e** Schematic representation for gene editing using the Cas12a nuclease which generates sticky-end DSB. **f** HDR efficiencies of HDR-boosting modular ssDNA donors and canonical ssDNA donors cooperated with Cas12a at the *RNF2* site in HEK 293T cells. 32 pmol Cas12a nuclease, 37.5 pmol gRNA and 6 pmol ssDNA donors corresponded to 2×10^5 cells. For all HDR efficiency-assessing experiments, HDR efficiency was measured three days after electroporation. Values and error bars reflect mean $\pm$ SD of $n = 3$ (**d**) independent electroporation replicates. Values reflect $n = 2$ (**b**, **f**) independent electroporation replicates. The sequences of all gRNAs and ssDNA donors used are shown in Supplementary Data 1 and 2. Source data are provided as a Source Data file. Figure (**a**, **c**, **e**) Created with BioRender.com released under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International license.

gene editing efficiency. We addressed this hypothesis by introducing base replacement at both exogenous *BFP* and endogenous gene loci in K562 cells. The additional HDR efficiency induced by these modular donors was from 36.1% to 44.4% for *BFP*, from 80.9% to 83.3% for *FANCF*, and from 58.3% to 68% for *FRMD7* (Supplementary Fig. 12b, c),

suggesting the consistent compatibility of our HDR-boosting modular donors to HDRobust. However, when the HDRobust strategy yielded very high HDR efficiency, such as at the *FANCF* site, there was relatively limited room for the modular donors to further improve HDR efficiency.

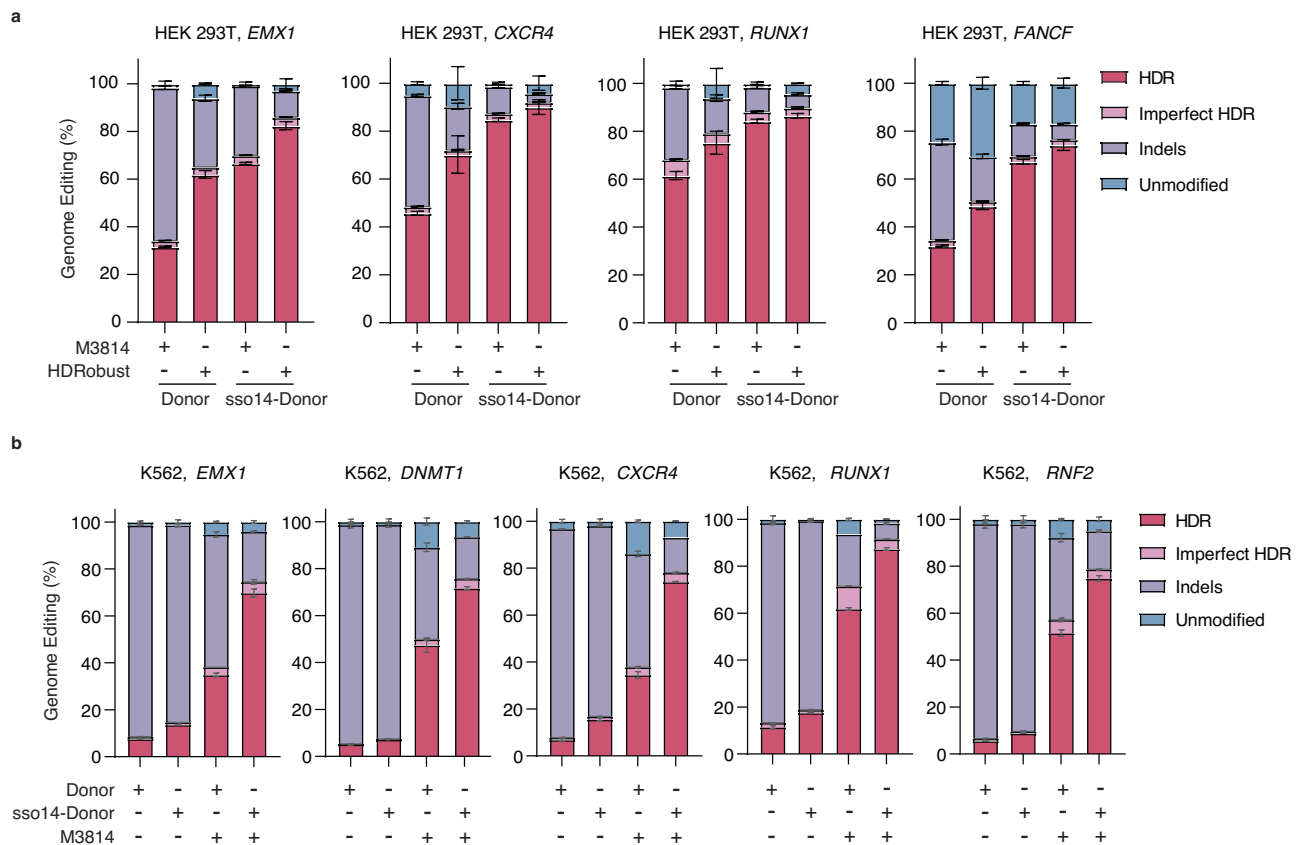


Fig. 6 | Enhanced potency of HDR-boosting modular donors with HDRobust or M3814. a Genome editing efficiencies at four endogenous loci in HEK 293T cells introduced by the indicated donors, along with transient end-joining inhibition by HDRobust (M3814 + POLQ siRNA mix) or M3814. **b** Genome editing efficiencies at five endogenous loci in K562 cells introduced by the indicated donors, with or without treatment by M3814. Electroporation was carried out using a 20 μ l mixture, containing 2×10^5 cells, 50.4 pmol Cas9 nuclease, 64 pmol gRNA and 40 pmol

ssDNA donors. When applicable, we added POLQ siRNA mix containing 32 pmol of POLQ siRNA predesigned pool (siRNAs 485, 1390, 1397 and 2460) and 64 pmol of POLQ siRNA 765. For transient NHEJ inhibition, 2 μ M M3814 was added for two days after electroporation, and editing efficiency was measured five days after electroporation. Values reflect $n = 2$ (a, b) independent electroporation replicates. The sequences of all gRNAs, ssDNA donors and POLQ siRNA used are shown in Supplementary Data 1, 2 and 6. Source data are provided as a Source Data file.

Considering M3814 remains the dominant component in HDRobust⁹, we investigated whether our modular ssDNA donor combined with M3814 can achieve comparable HDR efficiency while simplifying the components. Although HDR-boosting modular donors combined with HDRobust continued to provide the highest HDR efficiency (90.03%), combining HDR-boosting modular donors with only M3814 achieved higher HDR efficiency than canonical donors with HDRobust, yielding HDR efficiencies of 55.8% for *BFP*, and ranged from 66.6% to 84.8% for endogenous loci (Fig. 6a and Supplementary Fig. 12d). We further tested this combined strategy at five endogenous gene loci in K562 cells. In line with the results in HEK 293T cells, this combination achieved HDR efficiencies ranging from 69.7% to 87.3% (Fig. 6b), suggesting that M3814 robustly increases the potency of our HDR-boosting modules. Unexpectedly, this combination only achieved limited improvement in Cas12a-induced HDR, indicating different DNA repair pathways involved in repairing blunt DSB and sticky DSB (Supplementary Fig. 12e). Taken together, combining HDR-boosting modular ssDNA donor with either M3814 or HDRobust offers effective approaches for precise gene editing.

Discussion

The work described here defines a strategy that boosts the capabilities of ssDNA donors by incorporating functional sequence modules into the 5' ends of ssDNA donors. The HDR-boosting modules designed from the binding ODNs of the human RAD51 protein enhanced HDR efficiency mediated by Cas9 across multiple genomic loci and various

cell types. By combining with an inhibitor of NHEJ or the HDRobust strategy, the modular ssDNA donors achieved 66.62%–90.03% HDR efficiency, offering a simple, efficient and potentially safe method for enhancing precise gene editing.

Currently, three main approaches are employed to achieve precise gene editing: HDR, base editor (BE), and prime editor (PE)^{6,40}. Among these, HDR is versatile but relies on generating DSBs by programmable endonucleases, which can lead to unintended edits via end-joining pathways. Thus, enhancing HDR efficiency has been a major focus, with strategies falling into two categories: optimizing DNA donor design and directing the DSB repair pathway toward HDR.

DNA donor optimization often involves tethering ssDNA donors to Cas9 RNP via chemical modifications or additional proteins^{12,13,15}, the use of which is limited by its complexity and compatibility issues. By contrast, we developed HDR-boosting modules for ssDNA donors by leveraging the recruitment of endogenous RAD51, providing a chemical modification-free approach to enhance HDR efficiency. Although the role of RAD51 in SSTR remains controversial^{19,41}, it is consistently observed that RAD51 is recruited to DSB sites^{24,25,42}, which supports our approach. This active recruitment model is further supported by the higher editing efficiency achieved with a much lower concentration of modular donors compared to canonical ssDNA donors. As our HDR-boosting module and current Cas9 RNP-tethering strategies share a similar principle for enhancing the accessibility of ssDNA donors, they achieved comparable fold enhancement and HDR efficiency. However, nucleic acid delivery systems like nanoparticles and viral vectors favor

our modular DNA donors, which are readily co-delivered with nuclease mRNA or vectors. Together, by avoiding chemical or protein modifications that limit other Cas9 RNP-tethering strategies, HDR-boosting modules significantly advance donor delivery and efficiency.

Regulating the DSB repair pathway towards HDR by some molecular compounds has also been widely explored. Among them, the discovery of the HDR-improving effect by M3814 represents a significant breakthrough⁸. M3814, an inhibitor of DNA-PKcs, is one of the most potent HDR enhancers through blocking NHEJ^{43,44}. Its robust efficacy has been extensively studied in basic research in recent years^{44–46}. Notably, the group that identified M3814 as an HDR enhancer further developed the HDRobust approach⁹, which combined M3814 with a siRNA mix targeting Polθ to inhibit MMEJ. HDRobust approach has demonstrated superior capabilities in introducing higher HDR efficiency and outcome purity compared to base editors and prime editors⁹, making it a state-of-the-art strategy for precise gene editing. Given that M3814 remains the dominant component in HDRobust for improving absolute HDR efficiency in most editing sites⁹, we combined our HDR-boosting modular ssDNA donor with M3814. This combination surprisingly achieved comparable absolute HDR efficiency as HDRobust with canonical donors. For the sites potentially relying more on MMEJ repair, the addition of a siRNA mix targeting Polθ (HDRobust) yielded considerably higher HDR efficiency. In addition, this combination would also prevent potential off-target of Cas9 RNP. These findings not only underscore the compatibility of HDR-boosting modular donors but also highlight these combination strategies as efficient alternatives to the BE and PE gene editing techniques. Our study also indicates that DSB repair pathway selection and the recruitment of DNA templates are two major rate-limiting steps for achieving high HDR efficiency. Bridging DNA donor optimization and directing DSB repair pathways towards HDR would be a fundamental principle for developing efficient precise gene editing strategies in the future. As M3814 has entered phase I/II clinical trials⁸, the combination of HDR-boosting modular donor and M3814 holds great promise for therapeutic applications.

By targeting endogenous proteins via functional sequence modules, our chemical modification-free approach gets rid of the limitation of Cas9 RNP and can theoretically be compatible with any programmable nuclease, including the Cas12a verified in this study and other Cas12a variants^{47–49}. Besides enhancing DSB-induced HDR efficiency, HDR-boosting modules can also cooperate with paired Cas9 nickases to introduce precise genome editing with high specificity and eliminate the potential DSB cytotoxicity. Taking advantage of intrinsic and specific proteins involved in DNA nick-induced repair could further optimize modules for this approach. As ODIP-seq could capture both directly and indirectly bound ODNs, other proteins might also play a facilitating role in enhancing the efficacy of the HDR-boosting modular donors. Moreover, the Fanconi anemia pathway and RAD52 have recently been documented to engage in SSTR^{50,51}. It would be interesting to use these proteins to screen for biased binding sequences to facilitate the development of other HDR-boosting modules.

Conventionally, the study of protein functions has frequently entailed overexpressing mutant variants in knockout cells. However, this approach can result in unnatural protein levels, potentially introducing artifacts and leading to misinterpretations regarding protein function. Given the much higher frequency of HDR achievable compared to indels in commonly employed cell lines within this study, the combination approaches for directly and efficiently introducing endogenous point mutations within the native genomic context may simplify the manner in which gene function is explored in fundamental biological research. This can further offer precise and context-specific insights into gene functionalities, regulatory mechanisms, and disease processes.

Furthermore, the functional sequence module concept defined in this study may largely expand the functionality range of ssDNA donors,

which can serve as a platform for engineering customized and multi-functional donors in the future. According to the function of an ssDNA sequence as PAMP, ASO, and aptamer, connecting an immunostimulatory sequence to an ssDNA donor might enhance antigen presentation for a tumor vaccine⁵²; connecting a TLR9 suppression sequence to an ssDNA donor might evade immune responses for gene therapy⁵³; and connecting a specific ASO or aptamer might regulate gene expression or protein activity^{18,19}.

In summary, targeting endogenous protein RAD51 through an HDR-boosting module enables a chemical modification-free approach to improve gene editing efficiency. Thus, our study sheds light on the potential role of endogenous proteins in engineering DNA donors and provides a promising avenue for the development of powerful ssDNA donors for precise gene editing and translational applications.

Methods

Cell culture

HEK 293T, HeLa, and U2OS cells (Cell Resource Center, Peking Union Medical College) were cultured in Dulbecco's Modified Eagle Medium (DMEM) plus GlutaMax supplemented with 10% fetal bovine serum (FBS) (Gibco, Waltham, MA, USA). K562 cells (Cell Resource Center, Peking Union Medical College) were cultured in Roswell Park Memorial Institute (RPMI) 1640 medium supplemented with 10% FBS. hPB CD34+ cells (W-20210091, Sailybio, Shanghai, China) were cultured in Serum-Free Expansion Medium (SFEM) supplemented with hSCF (100 ng ml⁻¹), hFLT3 (100 ng ml⁻¹), hIL-3 (20 ng ml⁻¹), and hIL-6 (20 ng ml⁻¹). All cell types were passaged every 2–3 days, maintained below 80% confluence, and cultured at 37 °C with 5% CO₂. We confirmed that all cells tested negative for mycoplasma.

Lentivirus production for generating cell lines

To package the lentivirus and generate stable cell lines, 3 × 10⁶ HEK 293T cells were seeded in a 100 mm² dish in DMEM supplemented with 10% FBS. When the cells reached 70–80% confluence, we transfected them with 42 μl Lipofectamine 3000 (Thermo Fisher Scientific, Waltham, MA, USA), 10 μg lentivirus transfer plasmid, 3.5 μg pMD2.G (Addgene #12259), and 7.5 μg psPAX2 (Addgene #12260), following the manufacturer's protocol. Four hours after transfection, the medium was replaced with fresh DMEM supplemented with 10% FBS. The virus-containing supernatant was collected 36 h and 72 h post-transfection and centrifuged at 12,000 g, 4 °C for 10 min to remove cellular debris, filtered through a 0.22 μm polyvinylidene difluoride (PVDF) filter (Millipore, Burlington, MA, USA), and stored at –80 °C.

Construction of HEK 293T, and K562 cell lines with integrated BFP sequence

Lentivirus expressing a blue fluorescent protein (BFP) reporter construct under the EF1α promoter (Addgene, #71825) was produced from HEK 293T cells as described above. To stably integrate the BFP sequence, 6 × 10⁵ cells were infected with lentivirus at MOI of 0.3 in 6-well plates (Corning, NY, USA) containing DMEM supplemented with 10% FBS and 8 μg ml⁻¹ polybrene (Sigma-Aldrich, St. Louis, MO, USA). Two days after infection, single clones with BFP fluorescence were sorted using fluorescence-activated cell sorting (FACS) (Sony MA900, Tokyo, Japan). The BFP copy numbers integrated in genome DNA were determined by qPCR. Several single clones were pooled as the BFP cell model for evaluation of HDR efficiency.

Electroporation of different cell types

Cells were electroporated using the Neon Transfection System (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's protocol with 2 × 10⁵ cells, 18 pmol of Cas9 protein (Integrated DNA Technologies, San Diego, CA, USA), and 22 pmol of synthetic sgRNA (GenScript, Piscataway, NJ, USA). Electroporation on

Neon Transfection System (Thermo Fisher Scientific, Waltham, MA, USA) was carried out using 10 μ l Neon tips with the following parameters: 1,150 V, 20 ms, and 1 pulse for HEK 293T cells (parameters for other cells are available on the Thermo Fisher Scientific website). After electroporation, cells were cultured in 12-well plates (Corning, NY, USA) with corresponding media supplemented with 10% FBS. Three days later, cells with GFP signals were selected via fluorescence-activated cell sorting (FACS) to measure HDR efficiency.

For genome editing assay with M3814 inhibitor, electroporation was done using the Neon Transfection System (Thermo Fisher Scientific, Waltham, MA, USA) with a 20 μ l mixture, containing 2×10^5 cells, 50.4 pmol Cas9 nuclease (Integrated DNA Technologies, San Diego, CA, USA), 64 pmol gRNA (GenScript, Piscataway, NJ, USA) and 40 pmol ssDNA donors (Sangon Biotech, Shanghai, China). Where applicable, added 32 pmol of POLQ siRNA predesigned pool (containing siRNAs 485, 1390, 1397, and 2460) and 64 pmol of POLQ siRNA 765 (GenePharma, Suzhou, China). For transient NHEJ inhibition, 2 μ M M3814 (Selleck, China, S8586) was added for two days after electroporation, and editing efficiency was measured five days after electroporation.

Fluorescence-activated cell sorting (FACS)

To prepare for FACS, cells were separately washed with phosphate-buffered saline (PBS) 72 h after lentivirus infection or electroporation. They were then digested with 0.25% Trypsin-EDTA for 1 min, diluted to approximately 1×10^7 cells per ml with PBS, and filtered with a 40 μ m cell strainer cap (Corning, NY, USA). For cell sorting, infected cells were sorted for BFP (+) signals with equal mean fluorescence intensity. To analyze HDR efficiency, cells with different fluorescence signals were gated for further statistical analysis. The percentage of cells with GFP fluorescent signals denotes the HDR efficiency, and the percentage of cells converting from the BFP signal to double negative signals denotes the NHEJ efficiency. The cell gating strategy is shown in Supplementary Note 1.

ODN immunoprecipitation (ODIP)

HEK 293T cells were seeded in a 100 mm² dish containing DMEM supplemented with 10% FBS. When the cells reached approximately 90% confluency, the culture dishes were placed on ice and washed with ice-cold PBS. The PBS was drained, and ice-cold IP lysis buffer (Beyotime Biotechnology, Shanghai, China) was added to the culture dish at a ratio of 1 ml per 10^7 cells/100 mm² dish. Cells were then scraped off the dish using a cold plastic cell scraper and transferred into a pre-cooled microcentrifuge tube. After incubation at 4 °C for 30 min, the cell lysate was centrifuged at 12,000 g, 4 °C for 10 min to remove cellular debris. The cell lysate was then treated with 15 μ g 5' biotinylated ODN/ODN pool (24 nt, 200 ODNs) and incubated at 4 °C overnight with gentle agitation or rotation.

In parallel, antibody-protein G bead compounds were prepared in another microcentrifuge tube. The 5 μ g corresponding antibody {Anti-RAD51 antibody (abcam, ab133534, [EPR4030(3)]), Anti-Ku80 antibody (abcam, ab80592, [EPR3468]), Anti-Rad50 antibody (abcam, ab89, [13B3/2C6]), Anti-CtIP antibody (abcam, ab70163), Anti-53BP1 antibody (abcam, ab36823), Normal Rabbit IgG (Cell Signaling Technology, 2729 S)} was diluted into 500 μ l PBST with 20 μ l Protein G Dynabeads (Thermo Fisher Scientific, 10007D) and 100 μ g ml⁻¹ Salmon Sperm DNA (Thermo Fisher Scientific, 15632011). The mixture was incubated at 4 °C under rotary agitation for 4 h. Next, discard the supernatant, and the cell lysate incubated with ODNs was added to the beads, then the beads were incubated at 4 °C for 1 h under rotation. Following incubation, the supernatant was removed from the beads, and the protein of interest was specifically bound to the antibodies in the beads.

The 5' biotinylated ODNs bound with beads were incubated with streptavidin-HRP conjugate (Thermo Fisher Scientific, 89880D) and detected using a charged-coupled device (CCD) imaging instrument

(Tanon 5200, Shanghai, China). Besides, the ODN pool bound to the beads was digested using proteinase K (Servicebio, Wuhan, China) at 37 °C for 3 h and then extracted with phenol-chloroform for further ssDNA library preparation. The protein of interest was eluted in 40 μ l $2 \times$ SDS loading buffer for 10 min at 100 °C and detected using western blotting.

Single-stranded DNA library preparation for sequencing

A total of 112,764 24nt ODNs were generated based on missense single nucleotide polymorphism (SNP) sites in the ClinVar database (GRCh37/hg19), and 200 ODNs were randomly selected as the ODN screening pool for the ODIP-Seq assay. When the new ODN pool was obtained using the ODIP assay, the single-stranded DNA pool was converted into a library compatible with high-throughput sequencing using previously published methods⁵⁴. First, a 5'-phosphorylated adapter oligonucleotide was ligated to the 3' ends of the ODNs using the CircLigase II ssDNA Ligase (Lucigen, USA). Then the adapter-ligated molecules were immobilized on streptavidin beads (Thermo Fisher Scientific, 88817), and a 5'-tailed primer was used to copy the template strand with the Bst polymerase 2.0 (NEB, M0537L). After the removal of 3' overhangs via T4 DNA polymerase (Thermo Fisher Scientific, EP0062), a second adapter was added to the newly synthesized strands by blunt-end ligation with T4 DNA ligase (Thermo Fisher Scientific, EL0012). Then released the library molecules from the beads by heat denaturation and amplified the library for next-generation sequencing. The PCR amplification details are provided in the following "High-throughput amplicon sequencing of genomic DNA samples" subsection.

Biotin-ODN pulldown assay

HEK 293T cell lysates were obtained as mentioned above in the ODIP assay and divided into equal parts. Each 400 μ l aliquot of cell lysate was then treated with 4 μ g 5' biotinylated ODNs (24 nt) and incubated at 4 °C overnight with gentle agitation or rotation. Next, 20 μ l streptavidin magnetic beads (Thermo Fisher Scientific, 88817) were added to the cell lysate incubated with ODNs, and the samples were incubated at 4 °C for another 1–2 h under rotation. Following incubation, the supernatant was removed from the beads. The beads were washed with ice-cold PBST (0.05% Tween 20) for 3 times, with 10 min each. Finally, the proteins bound with ODNs were eluted in 20 μ l $2 \times$ SDS loading buffer for 10 min at 100 °C and detected using western blotting.

Silver staining

The proteins pulled down with biotin-labeled ssDNA donors were subjected to SDS-PAGE. After electrophoresis, the gel was washed with deionized water and covered with fix buffer (50% Methanol, 12% HAC, 0.05% Formaldehyde) for 2 h under slow rotation. Then the gel was washed with wash buffer (35% Ethanol) for 3 times, with 20 min each. After washing, the gel was covered with sensitizing buffer (0.02% Na₂S₂O₃) for 2 min, followed by washing with deionized water for 3 times, with 5 min each. Next, cover the gel with silver staining buffer (0.2% AgNO₃, 0.076% Formaldehyde) for 20 min, and keep away from the light. After staining, the gel was washed with deionized water twice, with 1 min each. Cover the gel with the developing buffer (6% Na₂CO₃, 0.05% Formaldehyde, 0.0004% Na₂S₂O₃), and stop staining with stopping buffer (50% Methanol, 12% HAC).

Mass spectrometry

Upon confirming the binding protein profile of the ssDNA donor via silver staining, an identical sample with the entire complement of binding proteins was subjected to mass spectrometry analysis using a Q-TOF platform (BGI Genomics, Shenzhen, China). Following the acquisition of the mass spectrometry data, the experimental group data was initially calibrated against the beads control to facilitate subsequent analysis. The mass spectrometry data is provided in the Source Data.

Genomic DNA extraction

HEK 293T, HeLa, U2OS, K562, and hPB CD34+ cells were cultured for 72 h after electroporation. The cells were washed with PBS and lysed with lysis buffer (20 mM Tris-HCl [pH 7.4], 1 mM CaCl₂) containing 800 units μl⁻¹ of proteinase K (Servicebio, Wuhan, China) at 37 °C for 3–4 h, followed by enzyme inactivation at 100 °C for 10 min. The genomic DNA was then isolated using phenol-chloroform and precipitated with absolute ethanol.

High-throughput amplicon sequencing of genomic DNA samples

Genomic sites were amplified from genomic DNA samples and sequenced using an Illumina NovaSeq platform (Illumina, San Diego, CA, USA). Briefly, an initial PCR step (PCR1) was used to amplify the target genomic sequence using primers containing Illumina forward and reverse adapters. In each 20 μl PCR1 mixture, 0.4 μM of each forward and reverse primer, 1 μl of genomic DNA extract (200 ng), 10 μl of 2 × Phanta Max Buffer, and 0.4 μl of Phanta Max Super-Fidelity DNA Polymerase (Vazyme, P505-d1, Nanjing, China) were used. The PCR1 conditions were as follows: 95 °C for 5 min, followed by 18 cycles of 95 °C for 15 s, 58 °C for 20 s and 72 °C for 90 s, and a final 72 °C extension for 5 min. A list of primers used for PCR1 reactions and the PCR1 amplicon sequences are provided in Supplementary Data 4 and 5. The subsequent PCR step (PCR2) added the unique i7 and i5 Illumina barcode combinations to both ends of the PCR1 DNA fragment to facilitate sample demultiplexing. In this step, 50 μl of a given PCR2 mixture contained 0.4 μM of each barcoding primer, 1 μl PCR1 product, 25 μl of 2 × Phanta Max Buffer, and 1 μl of Phanta Max Super-Fidelity DNA Polymerase (Vazyme, P505-d1). The barcoding PCR2 was carried out as follows: 95 °C for 5 min, followed by 22 cycles of 95 °C for 15 s, 58 °C for 20 s and 72 °C for 20 s, and a final 72 °C extension for 5 min. The PCR2 products were purified using a GeneJET PCR Purification Kit (Thermo Fisher Scientific, K0701) following the manufacturer's instructions and sequenced on an Illumina NovaSeq platform.

Restriction fragment length polymorphism assay

Genomic DNA was extracted and amplified as previously described. One microgram of purified PCR product was digested overnight at 37 °C with 20 U of Hind III and resolved on an agarose gel.

Tn5-based genome-wide sequencing for integration and translocation

Genomic DNA was extracted from HEK 293T cells three days after electroporation. Single-stranded adapter was synthesized by Sangon Biotech (Shanghai, China) and annealed to form a double-stranded adapter. Then double-stranded adapter was incubated with Tn5 enzyme (Vazyme, S601) at 30 °C for 1 h. Under the action of adapter-coupled Tn5 enzyme at 55 °C for 10 min, the genomic DNA was tagged into fragments of 500–1500 bp with the same adapter ligated at both ends. The tagged DNA fragments were purified using a DNA purification kit (Thermo Fisher Scientific, K0701). A pair of primers, targeting bait sequence and adapter, respectively, were used for DNA fragment amplification through 12 cycles of PCR, followed by an additional 12 cycles of nested PCR with another paired primers. Finally, 1 μl of nested PCR product was deployed with a third PCR for 26 cycles to introduce barcode and Illumina Adapters for next-generation sequencing (Read 1 for Tn5-adapter end and Read 2 for bait primer end). Sequencing was performed by Illumina HiSeq (PE150). Adapter and primer sequences are listed in the Supplementary Data 4.

After sequencing, raw reads were first processed with Trimmomatic to trim adapters and remove low-quality reads. The paired clean reads were filtered with the sequence from bait genomic regions. Since the unknown integrated or translocated sequences were captured by the Tn5-adapter end (Fig. 4a), the filtered reads in Read 1 files were

mapped to the hg19 genome by Bowtie2. Given a very low proportion of off-target integration reads and translocation reads compared to the reads mapping to the bait region, reads were not demultiplexed for downstream analysis. Off-target integrations and translocation junctions were binned to 0.1 Mb regions. The regions containing more than 1 read were regarded as faithful translocation regions and plotted.

Off-target editing analysis

The Cas-OFFinder tool was used to predict potential off-target sites at the *FANCF* genomic locus. The predicted off-target sites were amplified for next-generation sequencing, following the method of “High-throughput amplicon sequencing of genomic DNA samples” described above.

Sequencing data analysis

Trim Galore (version 0.6.6) was used to remove adapters and quality-trim all reads. For ODIP-seq analysis, ODN sequences were extracted and aligned to the ODN pool reference files to obtain the read numbers for each ODN using caRools (version 0.83). To calculate the relative Transcripts Per Million (TPM), the TPM value of a specific ODN in an ODIP sample was divided by the TPM value of the ODN in an IgG ODIP sample. The heatmap analysis was performed based on the relative TPM values. For the analysis of amplicons of endogenous sites, HDR and NHEJ frequencies were computed using CRISPResso2 with the amplicon sequence, ssDNA donor sequence, and sgRNA as inputs. The HDR and NHEJ frequencies were calculated by dividing the number of reads with HDR and NHEJ events by the total number of reads.

Statistics and reproducibility

No statistical method was used to predetermine sample size. Information concerning reproducibility for the experiments in this study are given in the corresponding figure legends. Sample sizes used in this study have been found to be sufficient for yielding reproducible results in mammalian cell gene editing experiments. No data were excluded from the analyses. The investigators were not blinded to allocation during experiments and outcome assessment. caRools (version 0.83) was used to analyze the ODIP-seq files. FlowJo was used to analyze the flow cytometry data. ImageJ was used for densitometry. CRISPResso2 (2.2.7) was used to analyze high-throughput sequencing files and quantify editing activity. Bowtie2 (2.4.1), trimmomatic (0.39), samtools (1.9), R (4.3.1), circlize (0.4.16), and pheatmap (1.0.12) were used to analyze NGS data for assess the safety of the modular ssDNA donors. Mean and standard deviations were calculated using GraphPad Prism 9.

Reporting summary

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

Data availability

Next-generation sequencing data and mass spectrum data generated in this study have been deposited to the China National Center for Bioinformation under accession code [PRJCA020579](https://www.ncbi.nlm.nih.gov/bioproject/PRJCA020579). All data generated or analysed during this study are included in this published article (and its supplementary information files). Source data are provided with this paper.

Code availability

All code used for processing library data is available on GitHub⁵⁵.

References

1. Awwad, S. W., Serrano-Benitez, A., Thomas, J. C., Gupta, V. & Jackson, S. P. Revolutionizing DNA repair research and cancer therapy with CRISPR–Cas screens. *Nat. Rev. Mol. Cell Biol.* **24**, 477–494 (2023).

2. Scully, R., Panday, A., Elango, R. & Willis, N. A. DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nat. Rev. Mol. Cell Biol.* **20**, 698–714 (2019).
3. Sfeir, A. & Symington, L. S. Microhomology-mediated end joining: A back-up survival mechanism or dedicated pathway? *Trends Biochem. Sci.* **40**, 701–714 (2015).
4. Bhargava, R., Onyango, D. O. & Stark, J. M. Regulation of single-strand annealing and its role in genome maintenance. *Trends Genet.* **32**, 566–575 (2016).
5. Yeh, C. D., Richardson, C. D. & Corn, J. E. Advances in genome editing through control of DNA repair pathways. *Nat. Cell Biol.* **21**, 1468–1478 (2019).
6. Zhang, X., Li, T., Ou, J., Huang, J. & Liang, P. Homology-based repair induced by CRISPR-Cas nucleases in mammalian embryo genome editing. *Protein Cell.* **13**, 316–335 (2021).
7. Jayavaradhan, R. et al. CRISPR-Cas9 fusion to dominant-negative 53BP1 enhances HDR and inhibits NHEJ specifically at Cas9 target sites. *Nat. Commun.* **10**, 2866 (2019).
8. Riesenberger, S. et al. Simultaneous precise editing of multiple genes in human cells. *Nucleic Acids Res.* **47**, e116 (2019).
9. Riesenberger, S. et al. Efficient high-precision homology-directed repair-dependent genome editing by HDRobust. *Nat. Methods* **20**, 1388–1399 (2023).
10. Roth, T. L. et al. Reprogramming human T cell function and specificity with non-viral genome targeting. *Nature* **559**, 405–409 (2018).
11. Ma, M. et al. Efficient generation of mice carrying homozygous double-flox alleles using the Cas9-Avidin/Biotin-donor DNA system. *Cell Res.* **27**, 578–581 (2017).
12. Savic, N. et al. Covalent linkage of the DNA repair template to the CRISPR-Cas9 nuclease enhances homology-directed repair. *Elife* **7**, e33761 (2018).
13. Aird, E. J., Lovendahl, K. N., St, M. A., Harris, R. S. & Gordon, W. R. Increasing Cas9-mediated homology-directed repair efficiency through covalent tethering of DNA repair template. *Commun. Biol.* **1**, 54 (2018).
14. Ling, X. et al. Improving the efficiency of precise genome editing with site-specific Cas9-oligonucleotide conjugates. *Sci. Adv.* **6**, eaaz51 (2020).
15. Carlson-Stevermer, J. et al. Assembly of CRISPR ribonucleoproteins with biotinylated oligonucleotides via an RNA aptamer for precise gene editing. *Nat. Commun.* **8**, 1711 (2017).
16. Lee, K. et al. Synthetically modified guide RNA and donor DNA are a versatile platform for CRISPR-Cas9 engineering. *Elife* **6**, e25312 (2017).
17. Zhang, P. et al. Schlafen 11 triggers innate immune responses through its ribonuclease activity upon detection of single-stranded DNA. *Sci. Immunol.* **9**, eadj5465 (2024).
18. Crooke, S. T., Baker, B. F., Crooke, R. M. & Liang, X. Antisense technology: an overview and prospectus. *Nat. Rev. Drug Discov.* **20**, 427–453 (2021).
19. Bouchard, P. R., Hutabarat, R. M. & Thompson, K. M. Discovery and development of therapeutic aptamers. *Annu. Rev. Pharmacol. Toxicol.* **50**, 237–257 (2010).
20. Ceccaldi, R., Rondinelli, B. & D Andrea, A. D. Repair pathway choices and consequences at the double-strand break. *Trends Cell Biol.* **26**, 52–64 (2016).
21. Hussmann, J. A. et al. Mapping the genetic landscape of DNA double-strand break repair. *Cell* **184**, 5653–5669 (2021).
22. Mulvey, C., Thur, B., Crawford, M. & Godovac-Zimmermann, J. Article Commentary: How many proteins are missed in quantitative proteomics based on MS/MS sequencing methods? *Proteom. Insights* **3**, PRI.S5882 (2010).
23. Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. & Corn, J. E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat. Biotechnol.* **34**, 339 (2016).
24. Aten, J. A. et al. Dynamics of DNA double-strand breaks revealed by clustering of damaged chromosome domains. *Science* **303**, 92–95 (2004).
25. Schrank, B. R. et al. Nuclear ARP2/3 drives DNA break clustering for homology-directed repair. *Nature* **559**, 61–66 (2018).
26. Nambiar, T. S., Baudrier, L., Billon, P. & Ciccica, A. CRISPR-based genome editing through the lens of DNA repair. *Mol. Cell.* **82**, 348–388 (2022).
27. Chen, F. et al. High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. *Nat. Methods* **8**, 753–755 (2011).
28. Lin, S., Staahl, B. T., Alla, R. K. & Doudna, J. A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife* **3**, e4766 (2014).
29. Frangoul, H. et al. CRISPR-Cas9 gene editing for sickle cell disease and β -thalassemia. *N. Engl. J. Med.* **384**, 252–260 (2020).
30. Huang, M. E. et al. C-to-G editing generates double-strand breaks causing deletion, transversion and translocation. *Nat. Cell Biol.* **26**, 294–304 (2024).
31. Liu, M. et al. Global detection of DNA repair outcomes induced by CRISPR-Cas9. *Nucleic Acids Res.* **49**, 8732–8742 (2021).
32. Kan, Y., Ruis, B., Takasugi, T. & Hendrickson, E. A. Mechanisms of precise genome editing using oligonucleotide donors. *Genome Res* **27**, 1099–1111 (2017).
33. Ran, F. A. et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154**, 1380–1389 (2013).
34. Davis, L. & Maizels, N. Homology-directed repair of DNA nicks via pathways distinct from canonical double-strand break repair. *Proc. Natl Acad. Sci.* **111**, E924 (2014).
35. Zetsche, B. et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759–771 (2015).
36. Dai, X. et al. One-step generation of modular CAR-T cells with AAV-Cpf1. *Nat. Methods* **16**, 247–254 (2019).
37. Ghanta, K. S. et al. 5'-Modifications improve potency and efficacy of DNA donors for precision genome editing. *Elife* **10**, e72216 (2021).
38. Gutierrez-Triana, J. A. et al. Efficient single-copy HDR by 5' modified long dsDNA donors. *Elife* **7**, e39468 (2018).
39. Renaud, J. et al. Improved genome editing efficiency and flexibility using modified oligonucleotides with TALEN and CRISPR-Cas9 nucleases. *Cell Rep.* **14**, 2263–2272 (2016).
40. Pacesa, M., Pelea, O. & Jinek, M. Past, present, and future of CRISPR genome editing technologies. *Cell* **187**, 1076–1100 (2024).
41. Gallagher, D. N. et al. A Rad51-independent pathway promotes single-strand template repair in gene editing. *Plos Genet* **16**, e1008689 (2020).
42. Mishra, A. P. et al. BRCA2-DSS1 interaction is dispensable for RAD51 recruitment at replication-induced and meiotic DNA double strand breaks. *Nat. Commun.* **13**, 1751 (2022).
43. Fu, Y. et al. Dynamics and competition of CRISPR-Cas9 ribonucleoproteins and AAV donor-mediated NHEJ, MMEJ and HDR editing. *Nucleic Acids Res.* **49**, 969–985 (2021).
44. Selvaraj, S. et al. High-efficiency transgene integration by homology-directed repair in human primary cells using DNA-PKcs inhibition. *Nat. Biotechnol.* **42**, 731–744 (2024).
45. Nyberg, W. A. et al. An evolved AAV variant enables efficient genetic engineering of murine T cells. *Cell* **186**, 446–460 (2023).
46. Chen, S., Chen, D., Liu, B. & Haisma, H. J. Modulating CRISPR/Cas9 genome-editing activity by small molecules. *Drug Discov. Today* **27**, 951–966 (2022).
47. Ma, E. et al. Improved genome editing by an engineered CRISPR-Cas12a. *Nucleic Acids Res.* **50**, 12689–12701 (2022).

48. Kleinstiver, B. P. et al. Engineered CRISPR–Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. *Nat. Biotechnol.* **37**, 276–282 (2019).
49. Zhang, L. et al. AsCas12a ultra nuclease facilitates the rapid generation of therapeutic cell medicines. *Nat. Commun.* **12**, 3908 (2021).
50. Richardson, C. D. et al. CRISPR–Cas9 genome editing in human cells occurs via the Fanconi anemia pathway. *Nat. Genet.* **50**, 1132–1139 (2018).
51. Paulsen, B. S. et al. Ectopic expression of RAD52 and dn53BP1 improves homology-directed repair during CRISPR–Cas9 genome editing. *Nat. Biomed. Eng.* **1**, 878–888 (2017).
52. Bo, Y. & Wang, H. Biomaterial-based in situ cancer vaccine. *Adv. Mater.* <https://doi.org/10.1002/adma.202210452> (2023).
53. Chan, Y. K. et al. Engineering adeno-associated viral vectors to evade innate immune and inflammatory responses. *Sci. Transl. Med.* **13**, eabd3438 (2021).
54. Gansauge, M., Meyer, M. & Single-stranded, D. N. A. library preparation for the sequencing of ancient or damaged DNA. *Nat. Protoc.* **8**, 737–748 (2013).
55. Jin, Y. et al. Enhancing homology-directed repair efficiency with HDR-boosting modular ssDNA donor. HDR-boosting_modular_donors. 10.5281/zenodo.12635855, 2024.

Acknowledgements

We thank all members of the De-Pei Liu laboratory for their helpful discussion about the project. This work was supported by grants from the National Key Research and Development Project of China (grant no. 2021YFA0804903, to D.L.), the National Natural Science Foundation of China (grant no. 92149305, to D.L.), the Haihe Laboratory of Cell Ecosystem Innovation Fund (grant no. HH22KYX0008, to D.L.), and the Chinese Academy of Medical Sciences Innovation Fund for Medical Sciences (grant no. 2021-I2M-1-016, 2022-I2M-2-001, and 2023-I2M-2-005, to D.L.).

Author contributions

Y.J. and P.Z. conceived the study and designed experiments; Y.J. performed the experimental studies; P.Z. conceptualized HDR-boosting modular ssDNA donor and analyzed the next-generation sequencing data; P.Z., L.L., X.Z., X.H., S.L., Z.L., Q.L., J.W., D.H., and Z.Z. assisted with experiments; Y.J. and P.Z. wrote the paper; D.L. and H.C. supervised the study.

Competing interests

D.L., Y.J., P.Z., and H.C. have submitted two patent applications to the China National Intellectual Property Administration pertaining to the HDR-boosting modules and the combination of HDR-boosting modular ssDNA donor and M3814 inhibitor aspects of this work (application number 2023108476157 and 2024107260446). The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-50788-x>.

Correspondence and requests for materials should be addressed to Hou-Zao Chen or De-Pei Liu.

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

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024