nature biotechnology

Article

https://doi.org/10.1038/s41587-022-01574-x

Precise cut-and-paste DNA insertion using engineered type V-K CRISPR-associated transposases

In the format provided by the authors and unedited

Supplementary Materials

Supplementary Tables

Supplementary Table 1: Plasmids and sequences of constructs used in this study

Supplementary Table 2: gRNA spacer sequences used in this study

Supplementary Table 3: Oligonucleotides and probes used in this study

Supplementary Table 4: Primary data

Note: *All Supplementary Tables are attached together as a single .xlsx file, separated by tabs*.

Supplementary Notes

Supplementary Figures and Legends

Supplementary References

Supplementary Notes:

Supplementary Note 1: Expanded discussion of Y2 ShHELIX results

While developing and characterizing ShHELIX, we also assessed whether the Y2 nAniI variant, previously shown to have a 9-fold higher affinity for its cognate target site[1], would enable a further increase in simple insertion product purity. With the Y2 ShHELIX construct, we observed a decrease in transformant colonies (**Sup. Fig. 3a**) when compared to ShCAST or non-Y2 ShHELIX (**Sup. Fig. 1a**). Moreover, this decrease varied with the spacing between the I-AniI site and LE/RE on pDonor, where a 14 bp spacing showed the highest number of colony-forming units (CFUs) (also aligning with the spacing giving the highest integration efficiency via ddPCR on plasmid and genomic targets). In combination with a similar observation when using a Lib4 I-AniI site (as shown in **Fig. 1k**), where the Lib4 I-AniI site was previously shown to increase wild type I-AniI affinity site by 5-fold[2], we recognized a potential correlation between the affinity of I-AniI for its target sequence and the number of colonies present on plates selecting for pShHELIX or pShCAST, pDonor and/or transposed product, and pTarget.

While further studies into the mechanism of HELIX will elucidate the basis of the decreased cell viability when using Y2-ShHELIX, we speculate that a combination of two phenomena may be occurring. First, the higher affinity of Y2 nAniI for its target, or when using nAniI with a Lib4 site, leads to an increased prevalence of DNA double-strand breaks (DSBs) on pDonor at early time points in the post-transformation recovery. In the absence of rapid and efficient cargo integration into pTarget, the AniI-caused DSBs result in a loss of Kanamycin resistance due to pDonor degradation prior to transposition. In this scenario, colony counts for different spacings on pDonor may correlate with higher or lower integration efficiencies. For example, for spacings where transposition is most efficient and rapid, the loss in CFUs is less striking because integration into pTarget occurs more rapidly than DSBs on pDonor. A second hypothesis is that the higher affinity of Y2 nAniI for its target, or when using nAniI with a Lib4 site, leads to an increased occurrence of DSBs on pDonor. Given the high copy number of pDonor in PIR1 cells, this could result in SOS response induction and cell death.

Supplementary Note 2: ShHELIX control experiments

While performing long-read sequencing of transposition products resulting from plasmid-targeting experiments, we included several control conditions. First, we performed experiments using a catalytically attenuated I-AniI variant (harboring K227M and Q171K mutations[3]) to create a 'dead' ShHELIX (dShHELIX). With dShHELIX, we observed a 1.8-fold decrease in cointegrate products compared to wild-type ShCAST (**Sup. Fig. 4a** and **Fig. 1i**, respectively). We hypothesize that this somewhat unexpected decrease in cointegrate products is the result of incomplete inactivation of I-AniI catalysis, which might lead to low-level 5' pDonor nicking (at a rate slower than nAniI-based ShHELIX). Indeed, the I-AniI Q171K variant has previously been shown to exhibit residual nicking activity on both DNA strands *in vitro*[3].

Secondly, we performed experiments using a pDonor variant that does not harbor I-AniI sites. In transformations with ShHELIX and this modified pDonor lacking I-AniI sites, we observed a 1.7-fold decrease in cointegrates relative to ShCAST (**Sup. Fig. 4a** and **Fig. 1i**, respectively). We hypothesize that this could be due to low-level I-AniI activity on sequences flanking the LE and RE (where tethering to TnsB induces energetically unfavorable interactions that would not occur in the absence of the fusion). A previous study that mutated each base in the I-AniI recognition sequence to all other bases revealed that specificity of nAniI is greatest across base pair positions ±3, 4, 5, and 6 in each half-site and least specific across bases -2 to +1 and bases at the outer edges of the recognition sequence[3]. From this data, a minimal approximate core sequence of 5'-GAGGNNNCTCTG-3' is necessary for I-AniI recognition, with decreased activity depending on the base substituted. While we could not identify an exact sequence match, we note that sequences similar to these core motifs occur on pDonor at 5'-GTGGNNNNGTCTA-3' (11 bp from the LE) and 5'-GAGGNNNCATTG-3' (13 bp from the RE), the latter being in an orientation that would give a nick on the same strand as TnsB (see next point). Low-level nicking on these flanking sequences at these degenerate I-AniI core sequences might lead to a slight increase in simple insertion product purity (as observed).

Thirdly, we performed experiments using 'flipped' I-AniI sites on pDonor oriented to confer a nick on the same strand as TnsB. In experiments using a flipped I-AniI site pDonor, we observed a 10-fold decrease in cointegrates with ShHELIX relative to ShCAST (**Sup. Fig. 4b**). We hypothesize that this reduction in cointegrates might be the result of an alternative transposition mechanism involving 5' flap cleavage of the gapped Shapiro intermediate (**Sup. Fig. 4c**).

Supplementary Note 3: Mechanistic implications of Cas12k-TnsC fusions

Recent structural studies have provided insight into the mechanism of ShCAST-mediated DNA insertion[4–6]. These studies suggest that TnsB recruitment to TniQ-nucleated TnsC filaments simulates filament disassembly, exposing the target site and inducing insertion at a coordinated distance from the sgRNA-Cas12k-DNA complex. Our experiments with fusions of Cas12k to a TnsC monomer in the context of ShCAST or ShHELIX (**Fig. 3**) are interesting given these proposed mechanisms, particularly regarding the role of TnsC filamentation in recruiting downstream transposition machinery. Additionally, since the extent of TnsC filament disassembly (or the footprint of TniQ alone or bound to TnsC) may define the insertion distance from bound DNA-bound Cas12k for canonical 4-component ShCAST, it is interesting that Cas12k-TnsC fusions (in the context of ShCAST and ShHELIX systems) enable targeted DNA insertion with the same insertion distance profiles as the canonical 4-component ShCAST and ShHELIX systems (**Sup Fig. 7**). We speculate that TnsC filamentation may still occur, despite Cas12k fusion, or that only a single TnsC subunit fused to Cas12k is sufficient to enable transposition. In the latter case, it is possible that TnsB-mediated depolymerization collapses TnsC filaments to a single monomer, which results in the fixed insertion distance profile observed for natural systems and would align with the identical profile observed for our monomer fusion. Alternatively, TnsC may not be involved in insertion distance determination, and a TniQ and TnsB defined insertion distance model may be more plausible. However, the molecular ruler mechanism of CASTs is still unclear. Furthermore, ShCAST our results revealed that a Cas12k-TniQ-TnsC fusion is functional (albeit with reduced activity) whereas a Cas12k-TnsC-TniQ fusion completely abolished activity (**Fig. 4b**). This observation may support the current model where Cas12k and TniQ must be able to directly interact[5]. Our results with Cas12k-TnsC and Cas12k-TniQ-TnsC fusions provide insight into the role of TnsC and TniQ in ShCAST-mediated transposition, motivating further studies to elucidate the transposition mechanism of both natural CASTs and engineered HELIX 2-, 3-, or 4-component systems.

Supplementary Note 4: Construction and characterization of N_7HELIX in human cell contexts

To construct N_7HELIX, a human codon optimized nicking variant of I-AniI was fused to N_7TnsB via an 18 amino acid XTEN linker. I-AniI sites were positioned 14bp from the LE and RE on pDonor in the correct orientation to confer a 5' nick, and the flanking sequences directly adjacent to the LE and RE were swapped for those of ShCAST (**Fig. 5a**). Although this donor flank configuration was most efficient for ShHELIX, it is possible that N_7-specific optimizations for N_7HELIX might yield higher integration efficiencies. To streamline N_7HELIX expression, we constructed a single all-in-one plasmid where all four HELIX components were driven by a single CMV promoter as previously described[7]. Specifically, NLS-Cas12k and TnsC as well as NLS-nAniI-TnsB and NLS-TniQ were linked by T2A sequences. Polypeptide pairs were separated by an EMCV internal ribosome entry site (IRES) (**Sup. Fig. 12c**). We also generated a modified version of the sgRNA (sgRNA2) with substitutions in several poly-T stretches within the scaffold of the wild-type sgRNA (which can serve as termination signal for the U6 promoter[8]) (**Sup. Fig. 12c**).

Recent work has demonstrated that host-encoded ribosomal protein S15 in bacteria is a bona fide component of type V-K CASTs, allosterically stimulating complex assembly at the Cas12k-bound target site[5]. Remarkably, the ShCAST sgRNA scaffold secondary structure to which S15 was found to be bound is strikingly similar to that of 16S rRNA (which S15 binds in its primary role in facilitating ribosomal complex assembly). Both *E. coli* S15 (EcS15) and *S. Hofmannii* S15 (ShS15) were previously shown to substantially enhance transposition *in vitro*[5]. Due to these observations, we generated expression plasmids for both N_7 ribosomal protein S15 (N_7S15) and EcS15 to determine if they could promote N_7CAST and N_7HELIX (**Fig. 5g**, **5h**, and **Sup. Fig. 13e**). We found that N_7S15 coexpression was required for N_7CAST and N_7HELIX integration in human cells (**Sup. Fig. 13e**), corroborating prior findings[5] that S15 is likely needed for optimal targeted integration and that it should be heterologously expressed when type V-K CASTs or HELIX is used in human cells. Under the conditions that we examined, we did not observe N_7CAST and N_7HELIX integration in human cells when EcS15 was coexpressed (**Sup. Fig. 13e**).

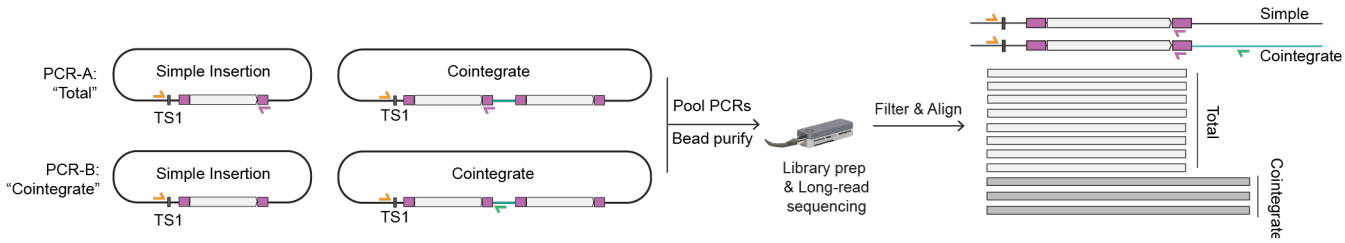
Despite detection of CAST- and HELIX-mediated transposition in human cells when expressing S15, overall insertion efficiency remained low for constructs and conditions tested. As expanded upon in our main text, discovering additional required host factors implicated in type V-K CAST function as well as screening for type V-K CAST orthologs that may be naturally suited for a human cell context will be needed. Directed evolution of CAST systems, particularly TnsB and Cas12k, and structure-guided engineering may enable more efficient integration on human genomic targets. Continued optimization of protein and sgRNA expression constructs and methods will also prove important given the complexity of these systems and the requirement to localize all components to the nucleus. Optimized component fusions may prove useful to help facilitate nuclear localization.

It should also be noted that the HELIX architectures may require optimization for each CAST ortholog. These optimizations include: spacing between the I-AniI site and LE/RE, linkers between nAniI and TnsB or between other components (if applicable), the identity of the LHE itself, and flanking sequences on the donor. System specific optimizations were not conducted for the other orthologs described in this study (AcCAST,

ShoCAST, and N_7CAST), as we designed and constructed N_7HELIX according to the optimal parameters from our ShHELIX/AcHELIX experiments. Therefore, ortholog-specific optimizations may enable more efficient HELIX-mediated human genome targeting.

<u>Supplementary Note 5</u>: Cointegrate characterization from experiments in HEK 293T cell lysates

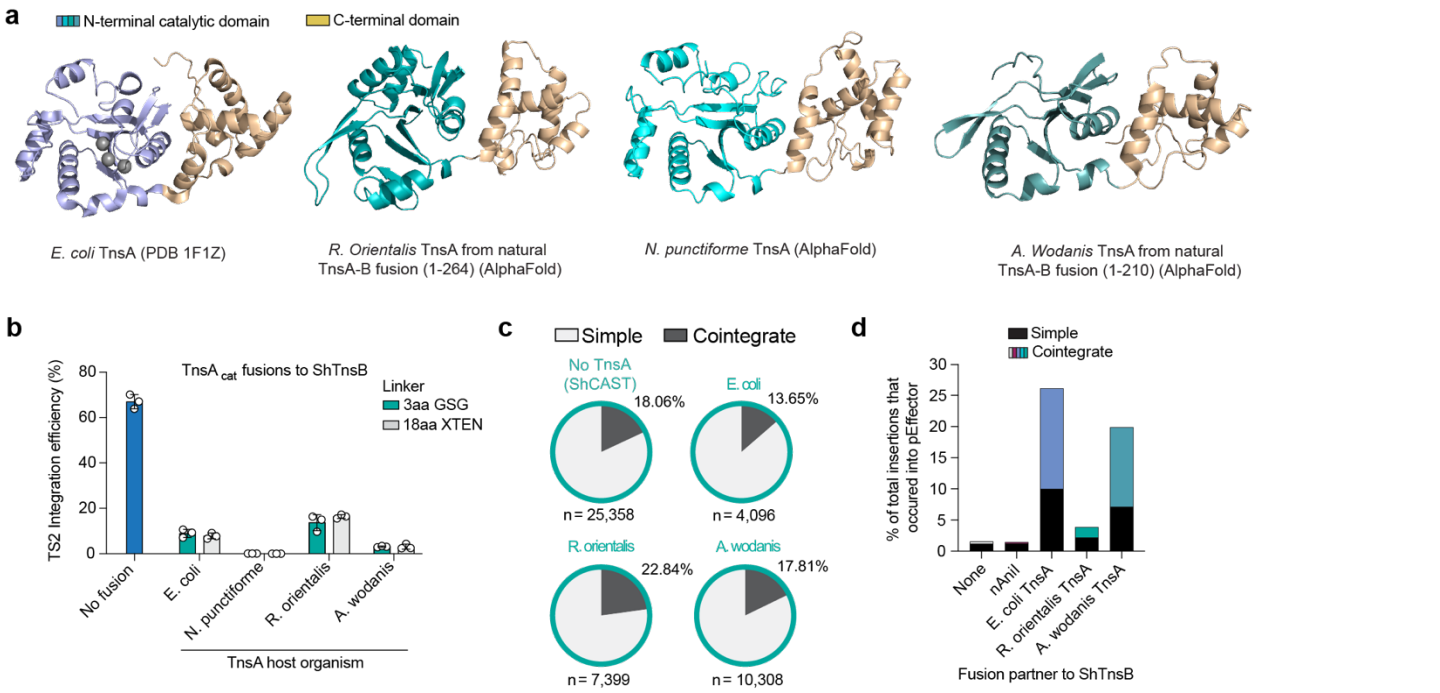
We explored the extensibility of HELIX to reduce cointegrates relative to its canonical CAST in human cell contexts. Due to low efficiency transposition in human lysates with the constructs and conditions that we examined, the enrichment process that we utilized for bacterial plasmid-targeting experiments was not feasible or applicable for experiments conducted in human lysate. Therefore, we opted to utilize a PCR-based enrichment strategy from the lysate reaction to quantify the approximate proportion of simple insertions to cointegrate products (see diagram below). Two separate 20-cycle PCRs each using an identical volume of terminated lysate reaction as template were conducted that differed only by the sequence of the downstream reverse primer. The PCRs sought to: (A) amplify from upstream of TS1 on pTarget to the edge of the RE on the inserted cargo (to approximate 'total' insertions), and (B) amplify from upstream of TS1 on pTarget (same 5' primer as first PCR reaction) to donor backbone near the edge of the RE. Both PCRs were performed for CAST and HELIX, the PCRs were combined and analyzed via long-read sequencing as described in methods. Reads from PCR-A represent "Total" insertions whereas reads from PCR-B represent "Cointegrate" insertions. The ratio of "Cointegrate" to "Total" insertions was used to estimate the relative proportion of cointegrates from total transposed product, albeit an approximate quantification and meant only to compare the relative differences between CAST and HELIX.



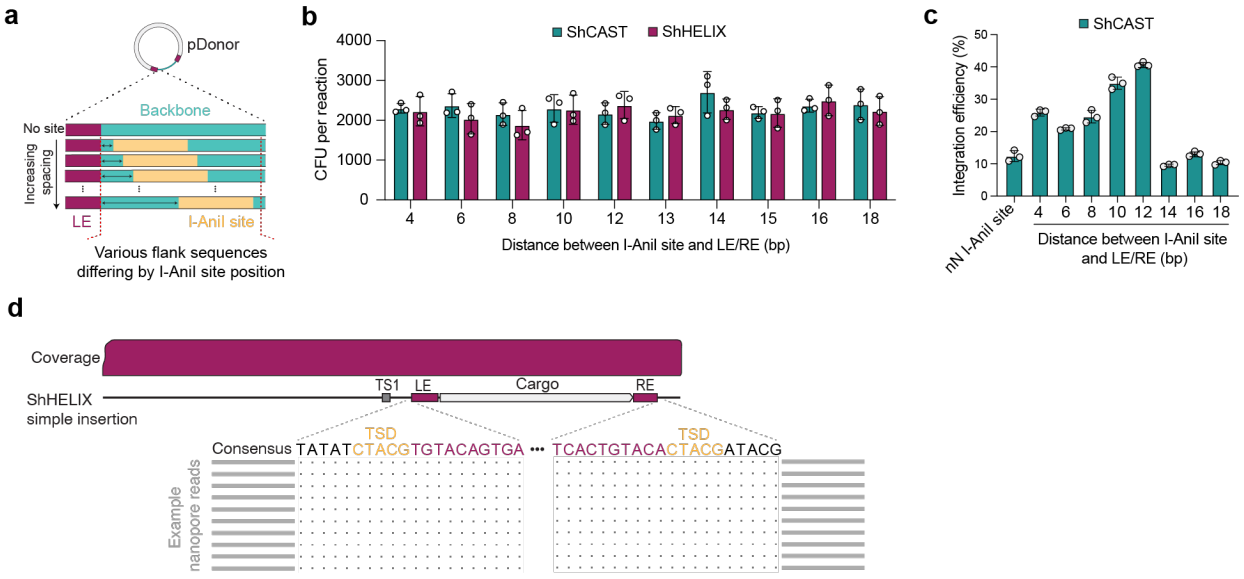
Supplementary Note 6: Script to calculate on-target integration from specificity analyses

```
import pandas as pd
import csv

def spec_analysis(file):

    '''This function calculates the proportion of integration events that
    occured on-target & provides a dataframe of integration coordinates
    to plot from'''


    # open txt file from bbtools analysis
    with open(file) as tsv:
        POS_lst = []
        SEQ_lst = []
        for line in csv.reader(tsv, dialect="excel-tab"):
            if '@' in line[0]:
                continue

            POS = int(line[3])
            SEQ = line[9]
            POS_lst += [POS]
            SEQ_lst += [SEQ]

    on_target = 0
    on_target_POS_lst = []

    # count insertions occuring within 55-75bp downstream of the PAM
    for i in POS_lst:

        # position 3867685 occurs right after PAM (GTT)
        if 3867610 < i < 3867630:
            on_target += 1
            on_target_POS_lst += [i] #can use to plot insertion distance

    # calculate and print on-target proportion
    fraction = on_target/len(POS_lst)
    print('ON-TARGET FRACTION: ' + str(fraction), on_target)
    print('TOTAL UNIQUE READS: ' + str(len(POS_lst)))

    df_pos_lst = pd.DataFrame(POS_lst)

    return df_pos_lst
```

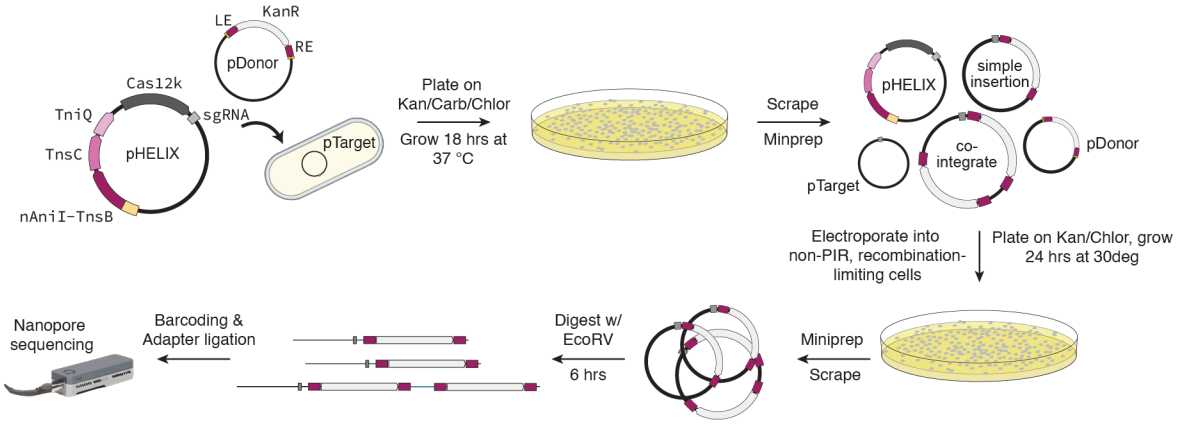
Supplementary Figures and Legends:



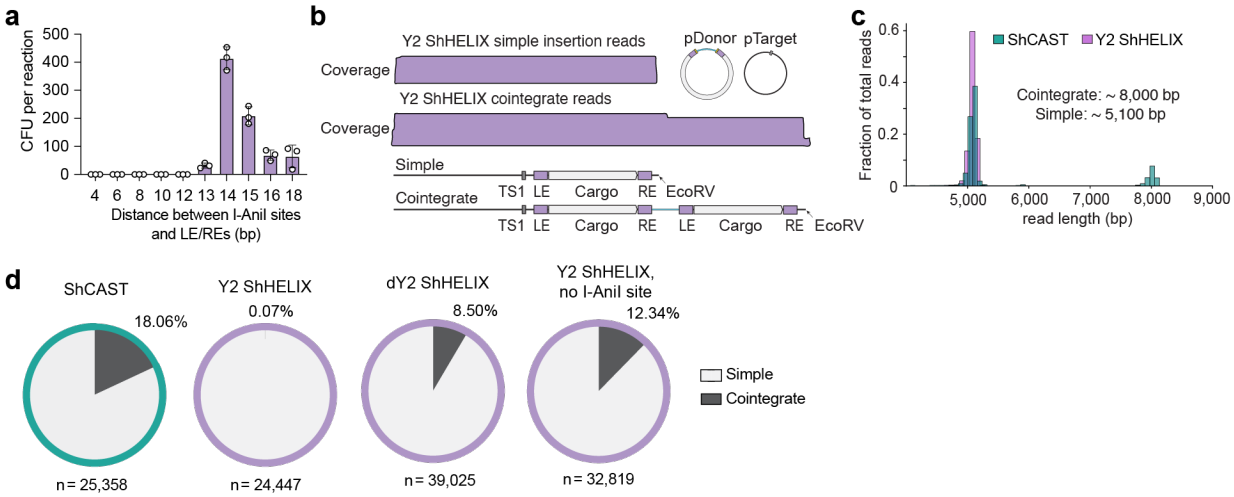
Supplementary Figure 1. Characterization of TnsA fusions to ShTnsB. a, Structures of various TnsA enzymes, either experimentally solved (*E. coli* TnsA; PDB 1F1Z*)* or computationally predicted via AlphaFold. **b,** Integration efficiencies when targeting genomic site TS2 using either ShCAST (no fusion) or variants containing fusions of TnsA and ShTnsB linked by either a short GSG or XTEN linker. Integration measured by ddPCR; mean, SD, and individual data points shown for *n* = 3 independent biological replicates. **c,** On-target cointegrate characterization as measured by long-read sequencing, following a Cas9-based target enrichment protocol. **d,** Proportion of total insertions that occur in the pEffector plasmid when using either no fusion (ShCAST), nAnil fusion (ShHELIX), or TnsA fusions.



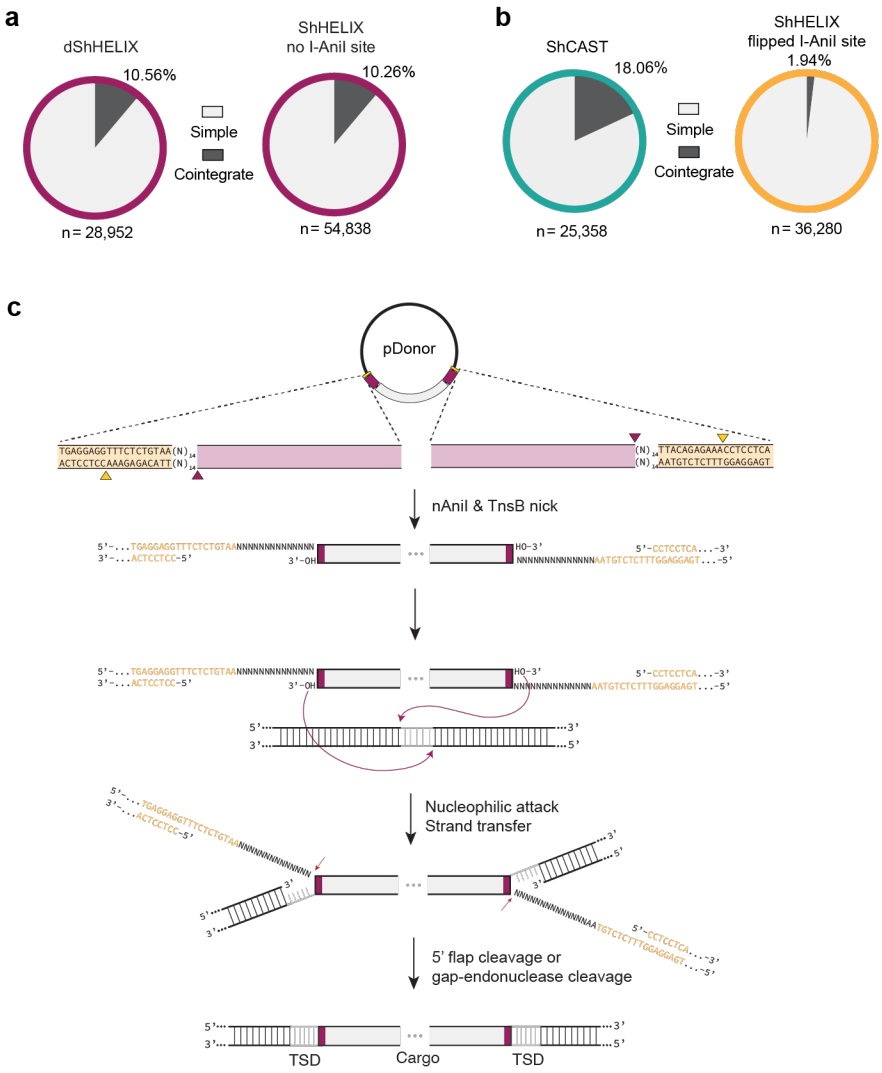
Supplementary Figure 2. Optimization and characterization of plasmid-targeting experiments. a, Schematic of donors bearing modified flank sequences with I-AniI sites positioned at various distances from the left and right transposon ends (LE/RE, respectively). **b**, Colony-forming units (CFUs) from transformations with ShCAST and ShHELIX plasmids targeting TS1 when using a series of pDonor plasmids bearing various spacings between the I-AniI sites and LE/RE. **c,** Integration efficiencies when using ShCAST targeting TS1 and a series of pDonors with different LE/RE flank sequences (corresponding to the ShHELIX pDonors bearing different spacings between the I-AniI sites and the LE/RE; see **panel a**), assessed via ddPCR. **d,** Alignment of ten exemplary reads bearing ShHELIX-mediated cargo integration 62 bp downstream of the PAM on pTarget. For **panels b** and **c**, mean, SD, and individual data points shown for *n* = 3 independent biological replicates. LE and RE, left and right transposon ends, respectively.



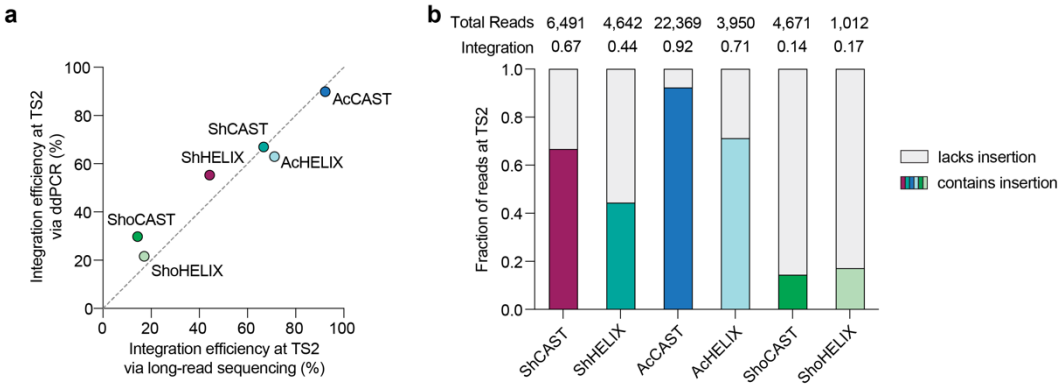
Supplementary Figure 3. Workflow for plasmid enrichment prior to long-read sequencing. Schematic of the protocol to enrich for transposed plasmid products to improve read-depth of intended products via long-read sequencing. sgRNA, single guide RNA; LE and RE, left and right transposon ends, respectively.



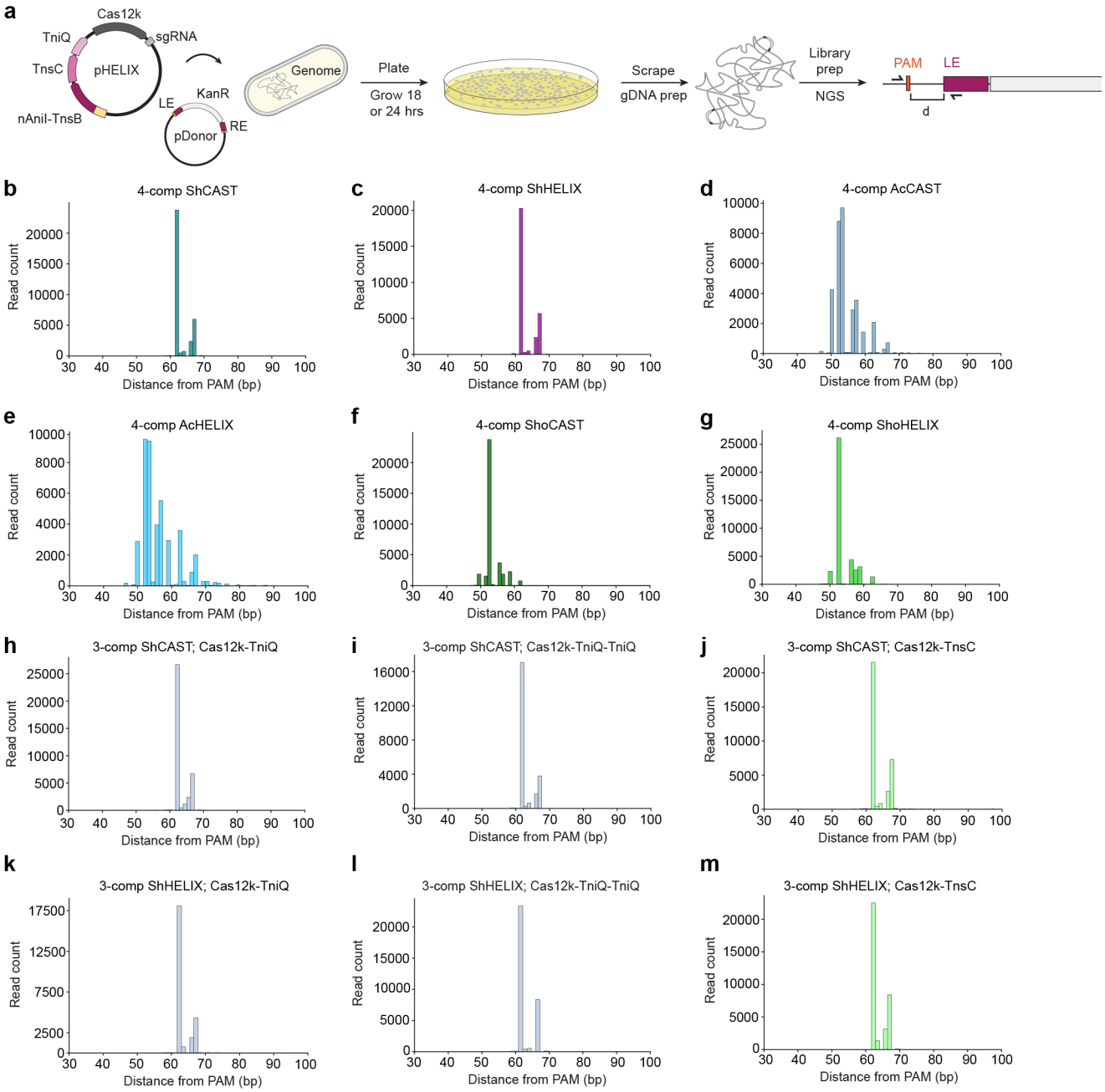
Supplementary Figure 4. Characterization of Y2 ShHELIX. a, Colony-forming units (CFUs) from transformations with Y2 ShHELIX plasmids targeting TS1 when using a series of pDonor plasmids bearing various spacings between the I-AniI sites and LE/RE. Mean, SD, and individual data points shown for *n* = 3 independent biological replicates. **b,** Coverage of expected insertion products into pTarget from long-read sequencing, displaying an exemplary subset simple insertion or cointegrate reads for Y2 ShHELIX. **c**, Read length distribution when using ShCAST and Y2 ShHELIX with a sgRNA targeting TS1 on pTarget. **d,** Comparison of simple insertion and cointegrate product proportions via long-read sequencing for various conditions using Y2-ShHELIX targeting TS1. LE and RE, left and right transposon ends, respectively.



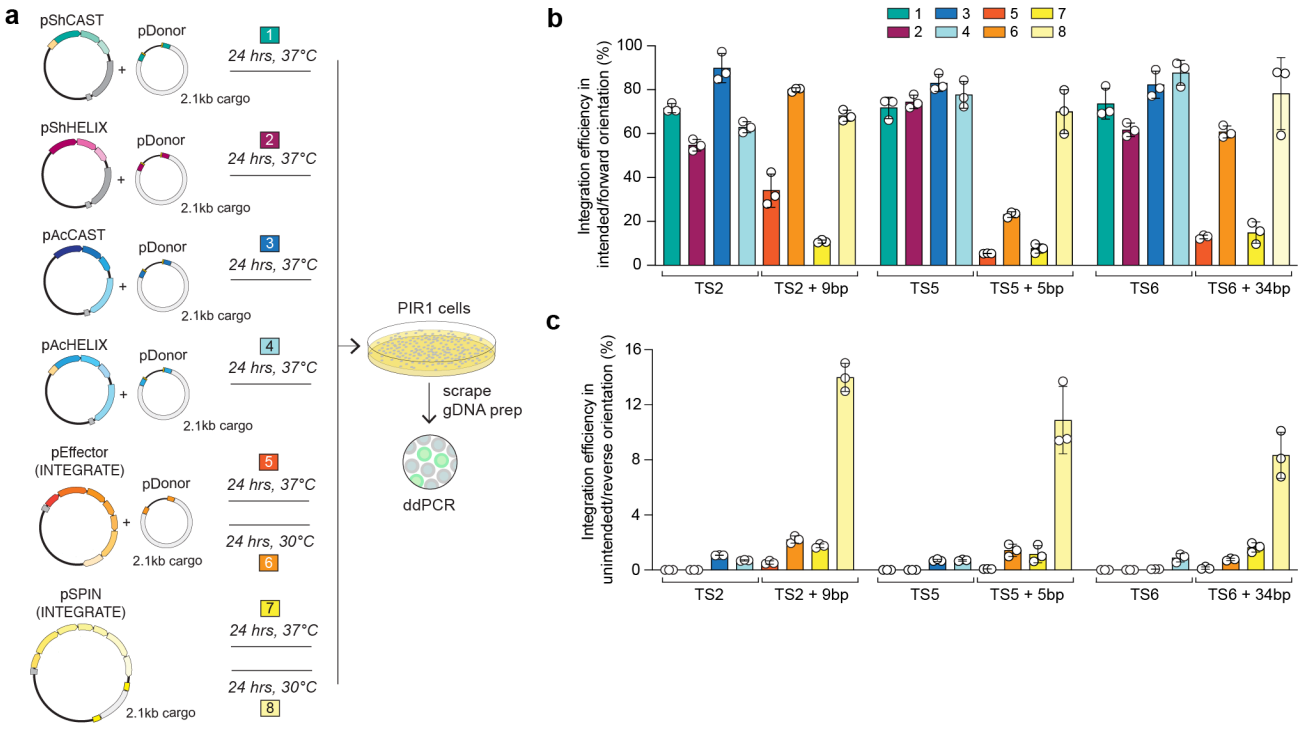
Supplementary Figure 5. ShHELIX control experiments. a, Comparison of simple insertion and cointegrate product proportions via long-read sequencing for a HELIX variant with a catalytically attenuated nAniI (dShHELIX) and when using HELIX with a pDonor without I-AniI sites. **b,** Comparison of simple insertion and cointegrate product proportions via long-read sequencing for ShCAST and ShHELIX when using a pDonor with flipped I-AniI sites that place the nAniI nicking sites on the same strand as the nick from TnsB. **c,** Potential alternative mechanism enabling simple insertion products when using a pDonor containing a flipped I-AniI site. TSD, target site duplication.



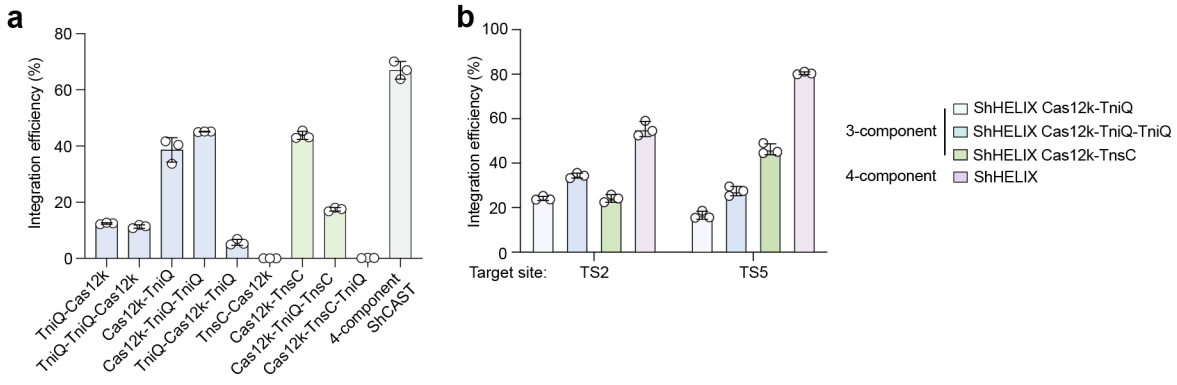
Supplementary Figure 6. Integration efficiency based on long-read sequencing. a, Comparison of integration efficiencies for each system as measured by ddPCR or by Cas9-enriched long-read sequencing. The dashed grey line denotes the diagonal (agreement between the two types of measurements). **b,** Integration efficiencies at TS2 when using CAST and HELIX systems, assessed via long-read sequencing. Stacked bars represent the fraction of Cas9-enriched target reads that lack or contain the cargo insertion. Integration (colored portion of each bar) represents the number of reads that contain the cargo insertion divided by the total number of targeted reads.



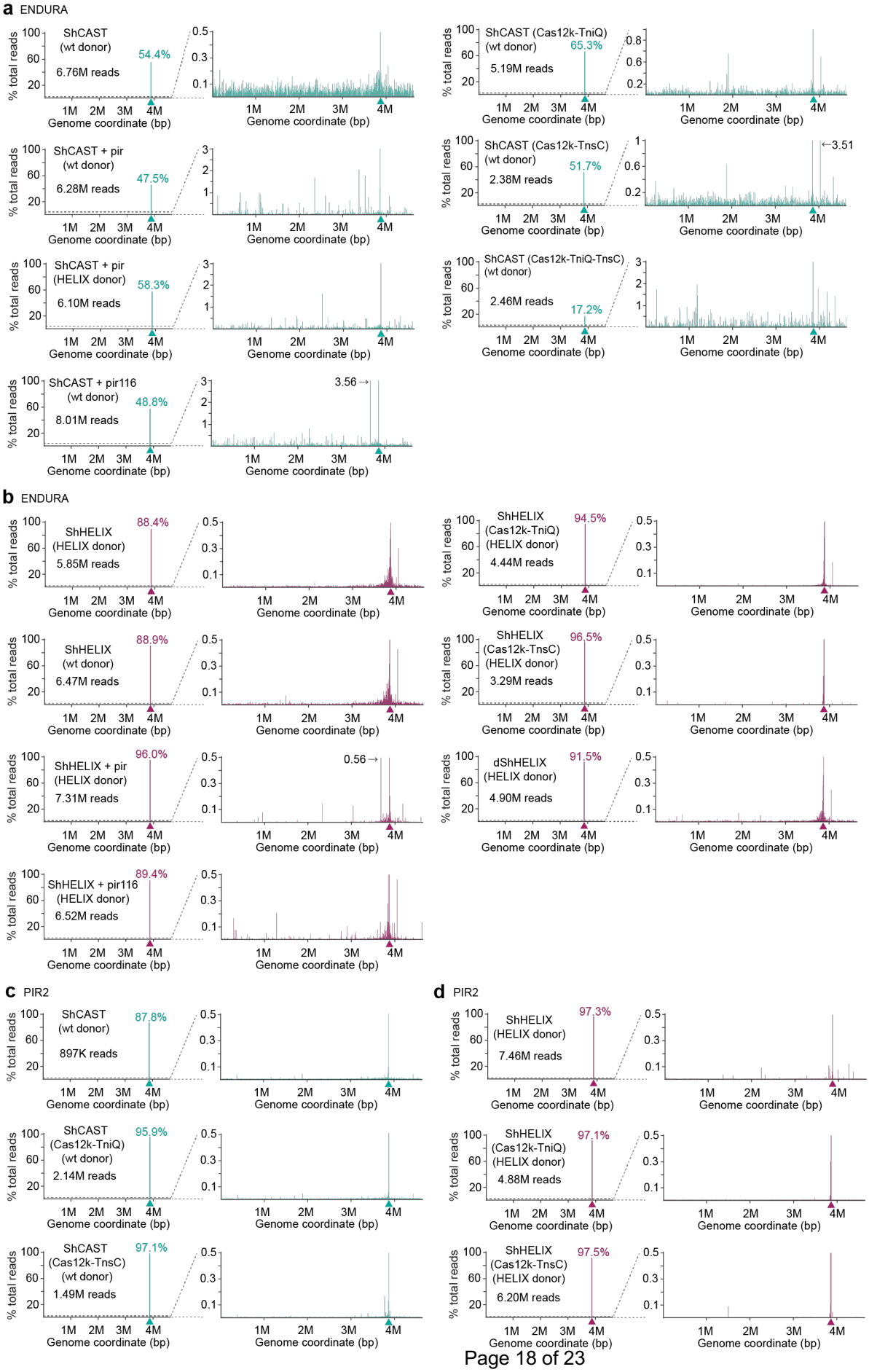
Supplementary Figure 7. Cargo insertion distance from the PAM. a, Schematic of the workflow to characterize PAM-to-LE insertion distances via next-generation targeted sequencing. PAM-to-LE insertion distance profiles for various CAST and HELIX constructs shown in **panels**: **b**, ShCAST (4-components); **c,** ShHELIX (4-components); **d,** AcCAST (4-components); **e,** AcHELIX (4-components); **f,** ShoCAST (4-components); **g,** ShoHELIX (4-components). **h,** ShCAST with Cas12k-TniQ (3-components); **i,** ShCAST with Cas12k-TniQ-TniQ (3-components); **j,** ShCAST with Cas12k-TnsC (3-components); **k**, ShHELIX with Cas12k-TniQ (3-components); **l,** ShHELIX with Cas12k-TniQ-TniQ (3-components); **m,** ShHELIX with Cas12k-TnsC (3-components); sgRNA, single guide RNA; PAM, protospacer adjacent motif; LE and RE, left and right transposon ends, respectively; NGS, next-generation sequencing.



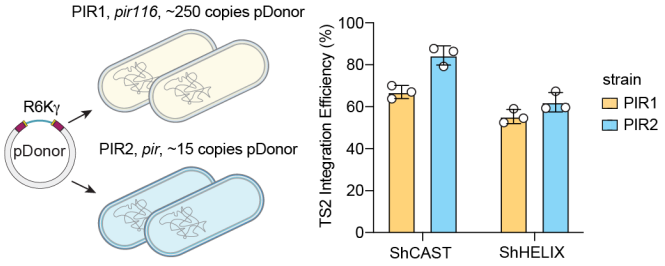
Supplementary Figure 8. Comparison of type I INTEGRATE and type V-K CAST and HELIX systems. a, Schematic of conditions and constructs tested, controlling for growth time (24 hrs), donor cargo size (2.1kb), approximate donor copy number (high copy), bacterial strain (PIR1), general target location (closest compatible PAMs near genomic target sites TS2, TS5, and TS6), and efficiency measurement method (ddPCR). **b,c,** Integration efficiencies of INTEGRATE, CAST, and HELIX in the intended forward orientation (**panel b**) or in the unintended reverse orientation (**panel c**). For **panels b** and **c**, mean, SD, and individual data points shown for *n* = 3 independent biological replicates.



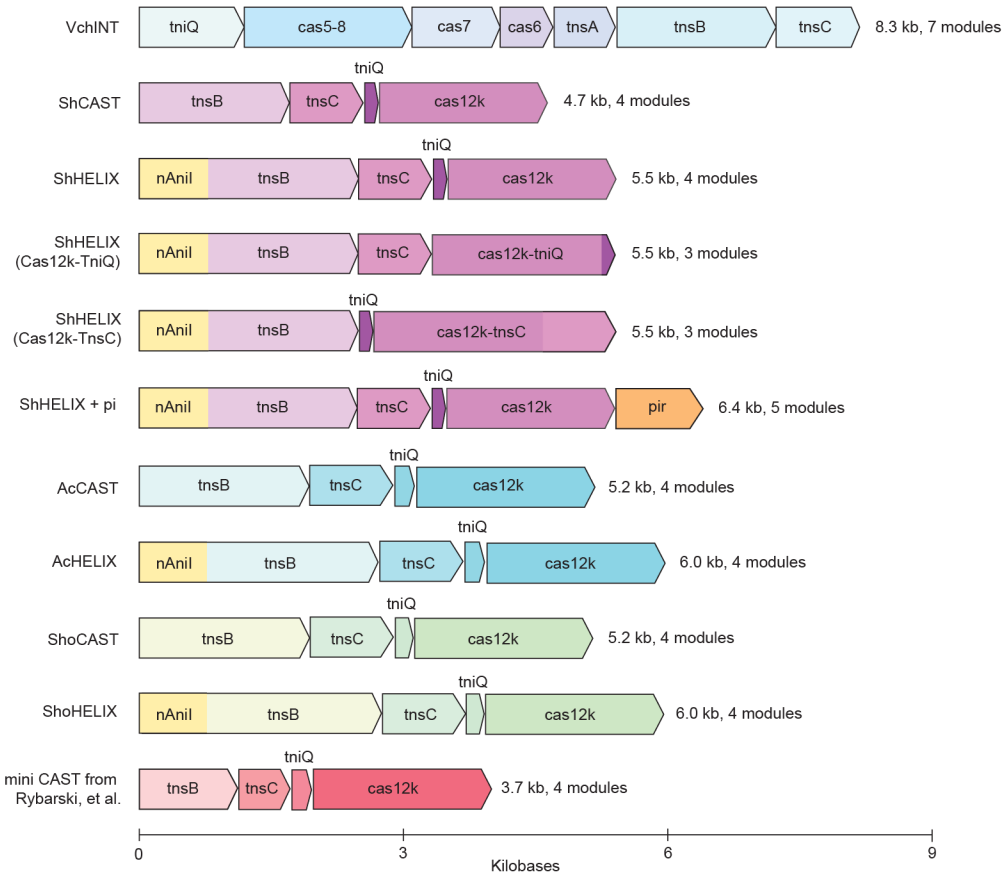
Supplementary Figure 9. Integration efficiencies for more minimal CAST and HELIX systems. a,b, Absolute integration efficiencies when targeting the genome at TS2 for 2-, 3-, or 4-component ShCASTs (**panel a**), and when targeting TS2 or TS5 for 3- and 4-component ShHELIX systems (**panel b**). For **both panels**, integration efficiencies were assessed via ddPCR and used to calculate relative integration as shown in **Fig. 3**; mean, SD, and individual data points shown for *n* = 3 independent biological replicates.


a ENDURA
ShCAST (wt donor) 6.76M reads 54.4%
ShCAST + pir (wt donor) 6.28M reads 47.5%
ShCAST + pir (HELIX donor) 6.10M reads 58.3%
ShCAST + pir116 (wt donor) 8.01M reads 48.8%
ShCAST (Cas12k-TniQ) (wt donor) 5.19M reads 65.3%
ShCAST (Cas12k-TnsC) (wt donor) 2.38M reads 51.7%
ShCAST (Cas12k-TniQ-TnsC) (wt donor) 2.46M reads 17.2%
b ENDURA
ShHELIX (HELIX donor) 5.85M reads 88.4%
ShHELIX (wt donor) 6.47M reads 88.9%
ShHELIX + pir (HELIX donor) 7.31M reads 96.0%
ShHELIX + pir116 (HELIX donor) 6.52M reads 89.4%
ShHELIX (Cas12k-TniQ) (HELIX donor) 4.44M reads 94.5%
ShHELIX (Cas12k-TnsC) (HELIX donor) 3.29M reads 96.5%
dShHELIX (HELIX donor) 4.90M reads 91.5%
c PIR2
ShCAST (wt donor) 897K reads 87.8%
ShCAST (Cas12k-TniQ) (wt donor) 2.14M reads 95.9%
ShCAST (Cas12k-TnsC) (wt donor) 1.49M reads 97.1%
d PIR2
ShHELIX (HELIX donor) 7.46M reads 97.3%
ShHELIX (Cas12k-TniQ) (HELIX donor) 4.88M reads 97.1%
ShHELIX (Cas12k-TnsC) (HELIX donor) 6.20M reads 97.5%
% total reads
Genome coordinate (bp)
1M 2M 3M 4M

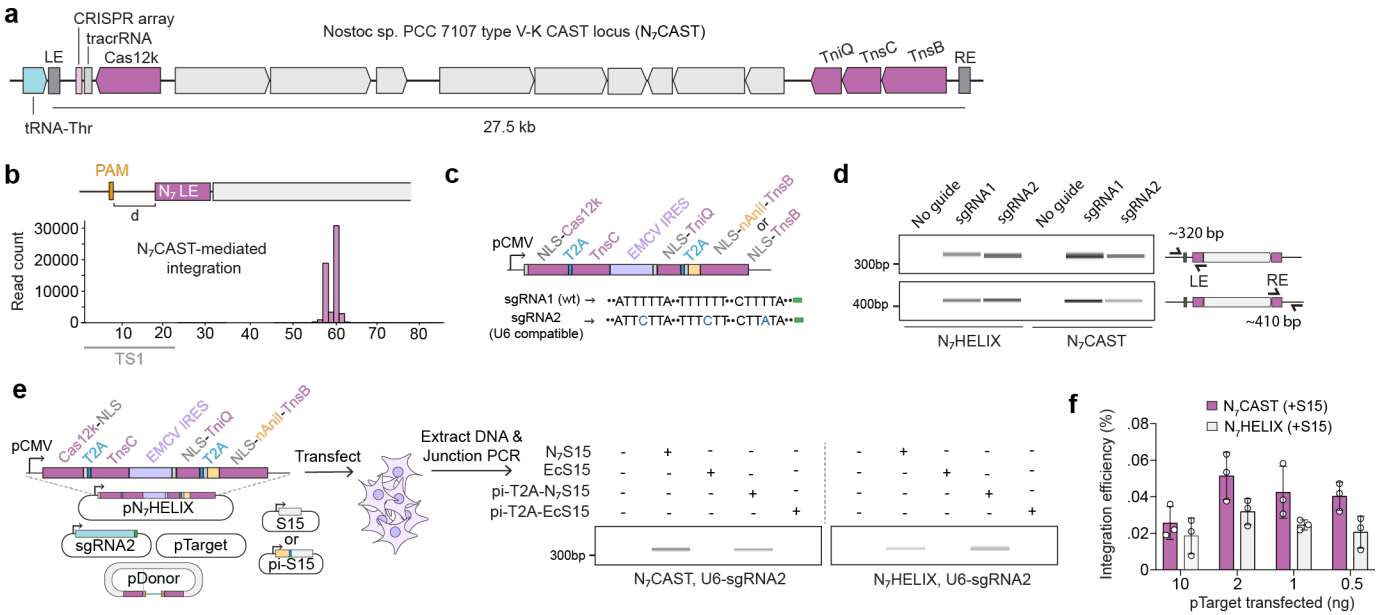
Supplementary Figure 10. Genome-wide integration profiles of ShCAST and ShHELIX systems. a-d, Integration site profiles from unbiased genome-wide insertion analysis of various CAST and HELIX constructs. The experiments were performed in Endura cells (**panels a** and **b**) or PIR2 cells (**panels c** and **d**), using various ShCAST configurations (**panels a** and **c**) or ShHELIX configurations (**panels b** and **d**) including different donor architectures, fusions to Cas12k, pi coexpression, or I-AniI variants.



Supplementary Figure 11. Influence of pDonor copy number and pi protein type on integration efficiency. Integration efficiencies using ShCAST and ShHELIX and an sgRNA targeting genomic site TS2 in two different bacterial strains that express either wild-type pi protein (*pir*) or a mutant copy-number mutant (*pir116*) (where PIR1 and PIR2 cells maintain pDonor at approximately 250 and 15 copies, respectively). Integration efficiencies assessed via ddPCR; mean, SD, and individual data points shown for $n = 3$ independent biological replicates. R6Kγ, origin of replication that requires the gene, *pir,* to replicate.



Supplementary Figure 12. Coding sequence and component number comparison of CAST and HELIX systems. Approximate sizes of coding sequences and number of protein subunits for prototypical type I and type V-K CASTs, HELIX systems developed in this study, as well as a recently described mini CAST from metagenomic mining[9]. nAniI, nicking I-AniI (K227M).



Supplementary Figure 13. Additional characterization of N_7CAST and N_7HELIX. a, Schematic of the genomic architecture of N_7CAST as found in Nostoc Sp. PCC7107 (identified by Strecker et al.[7]; not drawn to scale). **b,** PAM-to-LE insertion distance profile when using N_7CAST and an IVT sgRNA targeting TS1 on pTarget in lysate experiments, assessed by NGS. **c,** Schematic of all-in-one N_7CAST and N_7HELIX expression plasmids, and two versions of the sgRNA that either encode the canonical N_7 scaffold expressed from a U6 promoter (sgRNA1), or a derivative where poly-T stretches in the scaffold are substituted to be more compatible with transcription from the U6 promoter (sgRNA2). **d,** Junction PCRs when using N_7CAST or N_7HELIX with either IVT sgRNA1 or sgRNA2 targeting TS1 on pTarget in HEK 293T lysate experiments. **e,** Junction PCRs from HEK 293T cell-based plasmid-targeting experiments with or without N_7 or *E. coli* (Ec) S15 and pi proteins. For **panels d** and **e**, the Qiaxcel trace is a representative image from n = 2 independent replicates. **f,** Integration efficiencies of N_7CAST and N_7HELIX when transfecting various amounts of pTarget, as measured by ddPCR. Mean, SD, and individual data points shown for *n* = 3 independent biological replicates.

Supplementary References:

1. Takeuchi, R., Certo, M., Caprara, M. G., Scharenberg, A. M. & Stoddard, B. L. Optimization of in vivo activity of a bifunctional homing endonuclease and maturase reverses evolutionary degradation. *Nucleic Acids Res.* **37**, 877–890 (2009).
2. Scalley-Kim, M., McConnell-Smith, A. & Stoddard, B. L. Coevolution of a Homing Endonuclease and Its Host Target Sequence. *J. Mol. Biol.* **372**, 1305–1319 (2007).
3. McConnell Smith, A. *et al.* Generation of a nicking enzyme that stimulates site-specific gene conversion from the I-AniI LAGLIDADG homing endonuclease. *Proc. Natl. Acad. Sci.* **106**, 5099–5104 (2009).
4. Park, J.-U. *et al.* Structural basis for target site selection in RNA-guided DNA transposition systems. *Science* **373**, 768–774 (2021).
5. Schmitz, M., Querques, I., Oberli, S., Chanez, C. & Jinek, M. *Structural basis for RNA-mediated assembly of type V CRISPR-associated transposons*. http://biorxiv.org/lookup/doi/10.1101/2022.06.17.496590 (2022) doi:10.1101/2022.06.17.496590.
6. Park, J.-U., Tsai, A. W.-L., Chen, T. H., Peters, J. E. & Kellogg, E. H. Mechanistic details of CRISPR-associated transposon recruitment and integration revealed by cryo-EM. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2202590119 (2022).
7. Jonathan Strecker, Feng Zhang, Alim Ladha. Crispr-associated transposase systems and methods of use thereof. US2020/0190487A1
8. Gao, Z., Herrera-Carrillo, E. & Berkhout, B. Delineation of the Exact Transcription Termination Signal for Type 3 Polymerase III. *Mol. Ther. - Nucleic Acids* **10**, 36–44 (2018).
9. Rybarski, J. R., Hu, K., Hill, A. M., Wilke, C. O. & Finkelstein, I. J. Metagenomic discovery of CRISPR-associated transposons. *Proc. Natl. Acad. Sci.* **118**, e2112279118 (2021).