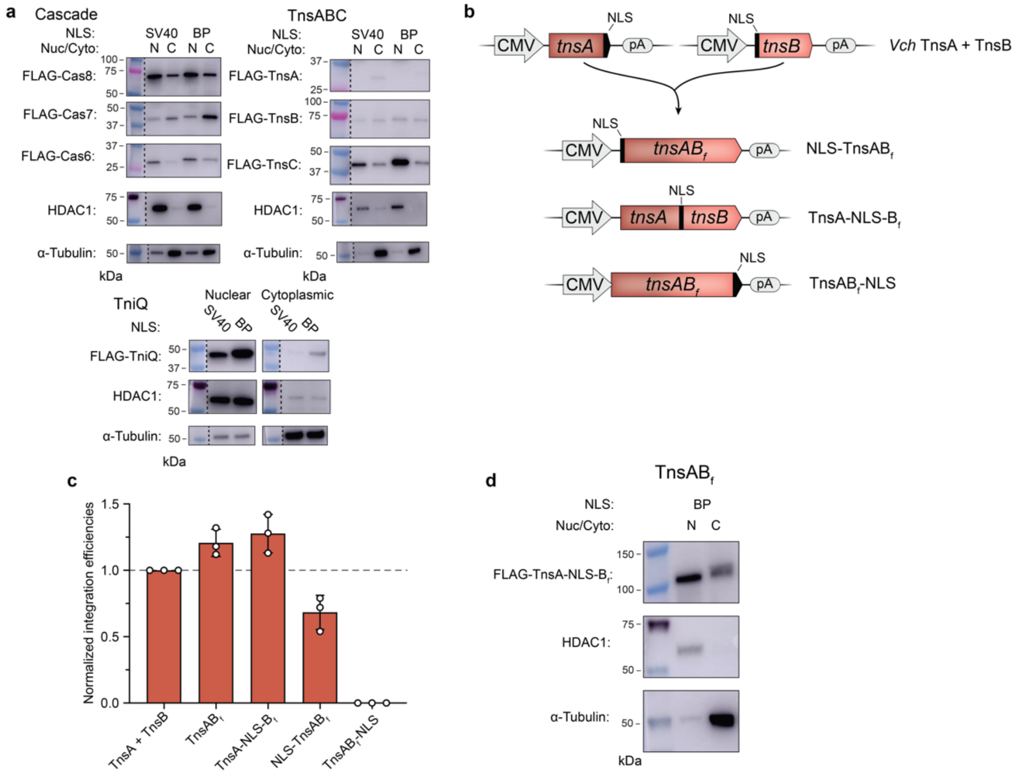
nature biotechnology

Article https://doi.org/10.1038/s41587-023-01748-1

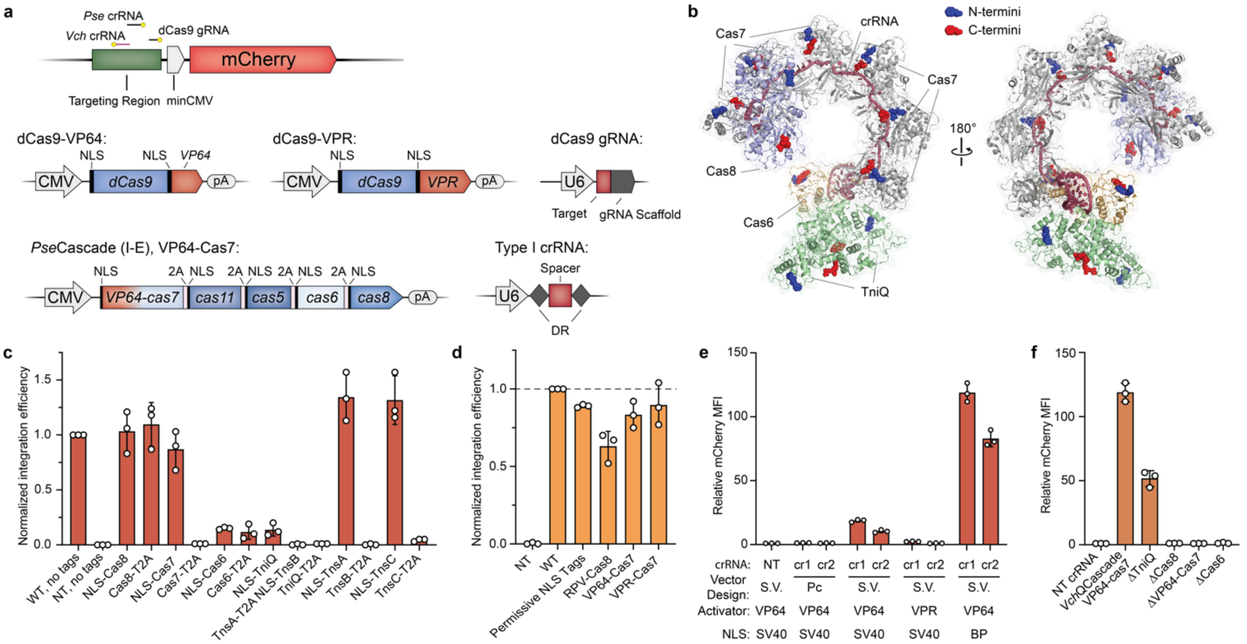
Targeted DNA integration in human cells without double-strand breaks using CRISPR-associated transposases

In the format provided by the
authors and unedited

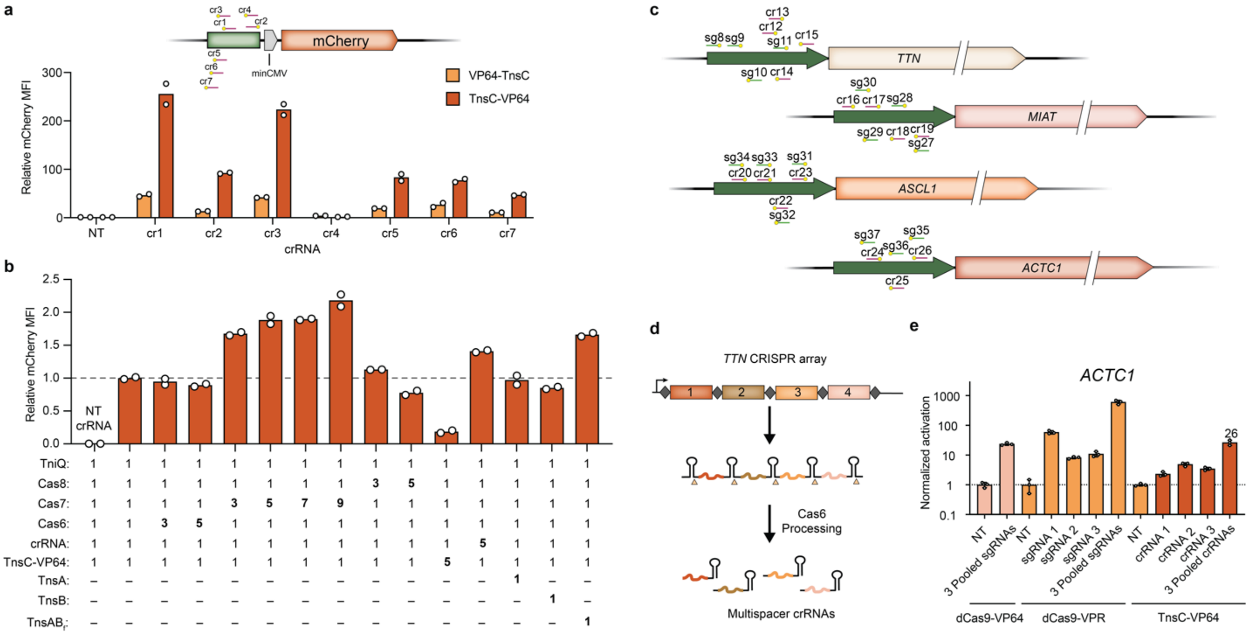
SUPPLEMENTARY INFORMATION



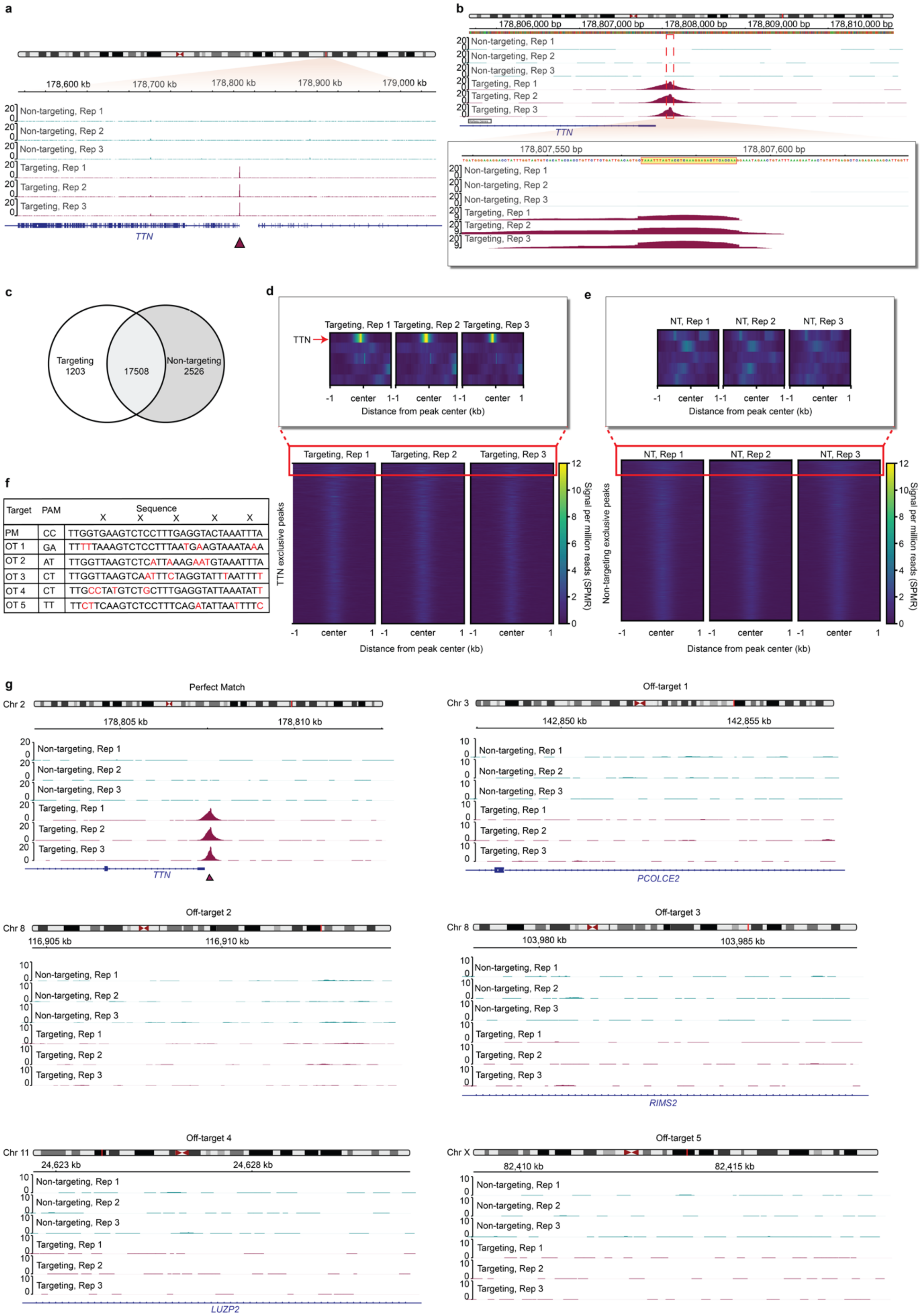
Supplementary Figure 1 | Improving expression and nuclear localization of *Vch*CAST components. a, Western blotting of various *Vch*CAST components using distinct nuclear localization signals (NLS). Each component was appended with a 3xFLAG epitope tag and NLS tag, and nuclear fractionation was performed to separate nuclear and cytoplasmic cellular proteins. Histone deacetylase 1 (HDAC1) and α-Tubulin were used as nuclear- and cytoplasmic-specific loading controls, respectively. Western blots were repeated in biological duplicate with similar results. **b,** Multiple fusion designs of TnsA and TnsB (TnsAB$_f$) were generated, with an NLS appended internally or at the N- or C-terminus. **c,** RNA-guided DNA integration activity was determined in *E. coli* with the indicated TnsAB$_f$ variants, as measured by qPCR. Data are shown as a mean ± s.d. for n = 3 biologically independent replicates. **d,** Western blotting of TnsAB$_f$ with internal NLS validated expression and nuclear localization. The observed band was at the expected size, with no evidence of degradation or internal cleavage. Western blots were repeated in biological duplicate with similar results. Uncropped western blots are shown in **Supplementary Figure 14.**



Supplementary Figure 2 | Optimization of *Vch*QCascade expression and transcriptional activation in human cells. a, Above, schematic of mCherry reporter plasmid for transcriptional activation assays. The location of sites targeted by Cas9 single-guide RNAs (sgRNA) and Cascade CRISPR RNAs (crRNA) are indicated. PAMs are marked with a yellow circle. Below, design of mammalian expression vectors encoding Cascade-based transcriptional activators from a Type I-E system (*Pse*Cascade)[1], alongside dCas9-VP64 and dCas9-VPR controls[2,3]. **b,** Depiction of *V. cholerae* TniQ-Cascade structure (PDB ID: 6PIF) showing the location of N- and C-termini in blue and red, respectively. All termini are solvent exposed and appear amenable to tagging. **c,** RNA-guided DNA integration activity in *E. coli* with the indicated NLS and/or 2A-tagged protein variants, measured by qPCR. Numerous tags have a deleterious effect. Data are normalized to the "WT no tags" condition, which resulted in a mean integration efficiency of 51 ± 8 %. **d,** RNA-guided DNA integration activity in *E. coli* with combined NLS and transcriptional activator fusions, as measured by qPCR. Fusing a VP64 or VPR transcriptional activator to the N-terminus of Cas7 exhibited the least deleterious effects on integration activity. Data are normalized to the "WT" condition, which resulted in a mean integration efficiency of 76.4% ± 4%. **e,** Strength of transcriptional activation across a set of distinct crRNAs ("cr#") targeting the mCherry reporter plasmid, as well as various activator-NLS constructs. Activation was measured using the reporter shown in **Supplementary Figure 2a** and measured by flow cytometry. S.V. indicates single vector design. Pc indicates polycistronic design of expression vectors as shown in **Supplementary Figure 2a. f,** Transcriptional activation by *Vch*QCascade utilizing a VP64-Cas7 fusion construct is strictly dependent on the presence of all Cascade components, as seen from the indicated dropout panel, but proceeds with ~50% activity in the absence of TniQ. Data in **c–f** are shown as mean ± s.d. for n = 3 biologically independent samples.

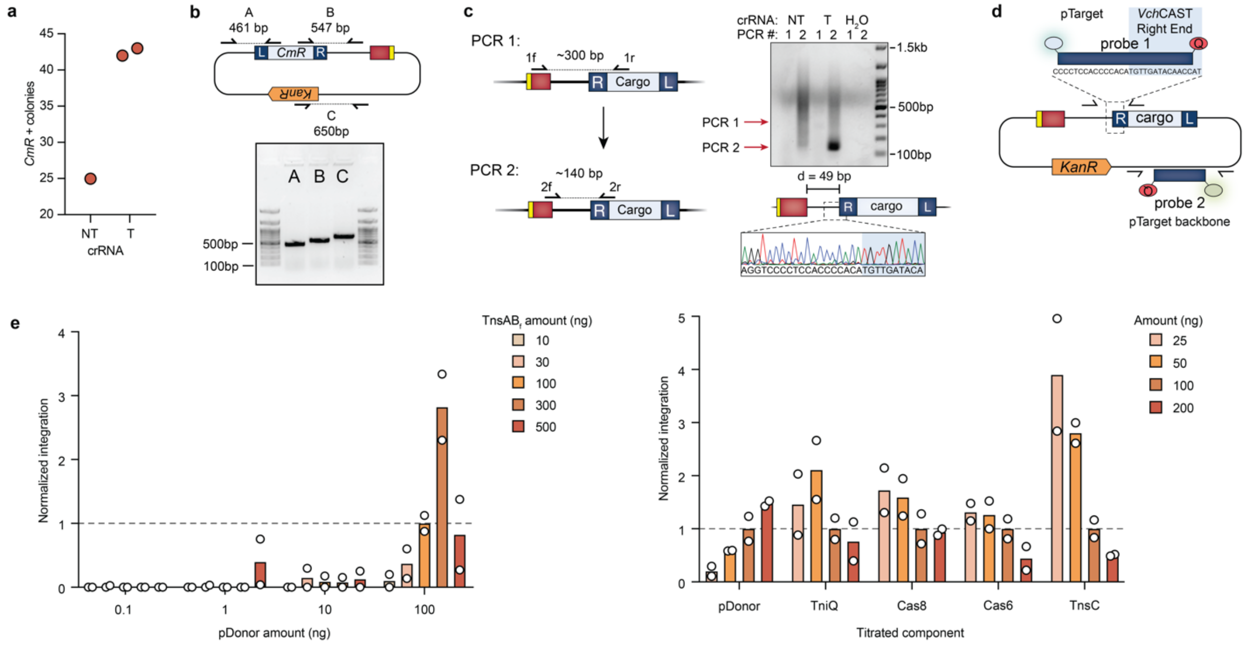


Supplementary Figure 3 | Optimization of TnsC-mediated transcriptional activation in human cells. a, Normalized mCherry fluorescence levels for the indicated experimental conditions, as measured by flow cytometry. VP64 was appended to TnsC at either the N- or C-terminus (VP64-TnsC or TnsC-VP64, respectively), and crRNAs ("cr#") were cloned to target various sites upstream of the mCherry gene (top). mCherry fluorescence levels were measured by flow cytometry and normalized to the non-targeting gRNA condition (bottom). **b,** Transcriptional activation is affected by titrating the relative levels of each expression plasmid, with numbers below the graph indicating the fold-change of each plasmid amount relative to the initial stoichiometric condition with a targeting crRNA (second bar from left). mCherry fluorescence levels were measured by flow cytometry. **c,** Schematic showing the position of crRNAs ("cr#") or sgRNAs (sg#) targeting each genomic locus for TnsC-mediated transcriptional activation for *Vch*CAST (maroon) and dCas9 *TTN* activation (green). **d,** Representative schematic of multispacer crRNAs used during TnsC-mediated genomic transcriptional activation. For *TTN*, *MIAT*, and *ASCL1*, the 4 individual spacer sequences used in individual or pooled crRNA conditions were expressed as one multispacer CRISPR array. The CRISPR array is processed by Cas6 after transfection into cells and enables programmable targeting of multiple copies of QCascade and TnsC to a target locus. **e,** Genomic transcriptional activation at the *ACTC1* locus as quantified by RT-qPCR. 3 distinct crRNAs or gRNAs were used for each condition. Data in **a** and **b** are shown as mean for n = 2 biologically independent samples. Data in **e,** are shown as the mean ± s.d. for n = 3 biologically independent samples.

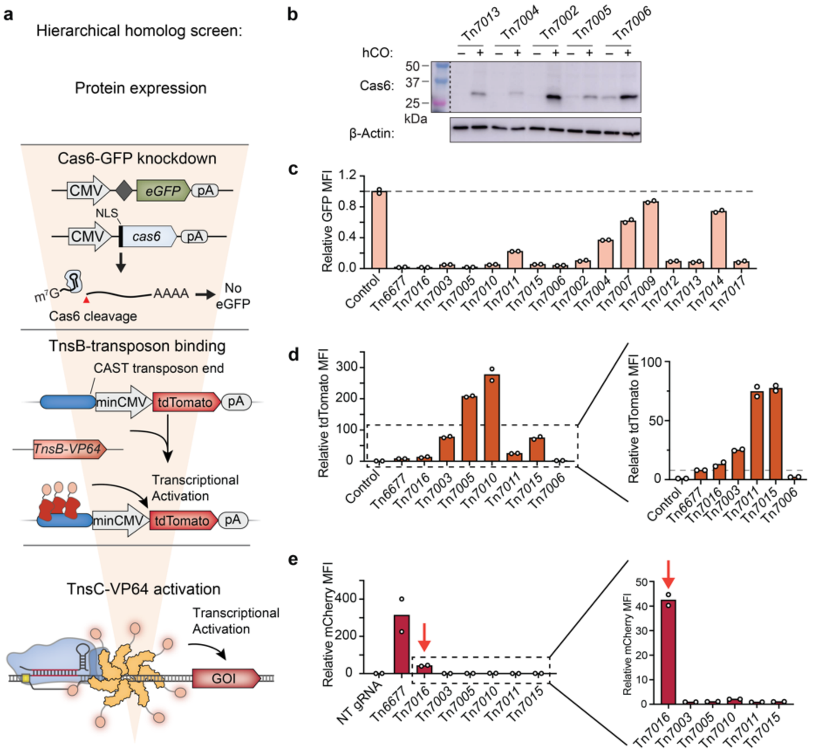


Target	PAM	Sequence
PM	CC	TTGGTGAAGTCTCCTTTGAGGTACTAAATTTA
OT 1	GA	TTTTTAAAGTCTCCTTTAATGAAGTAAATAAA
OT 2	AT	TTGGTTAAGTCTCATTAAAGAATGTAAATTTA
OT 3	CT	TTGGTTAAGTCAATTTCTAGGTATTTAATTTT
OT 4	CT	TTGCCTATGTCTGCTTTGAGGTATTAAATATT
OT 5	TT	TTCTTCAAGTCTCCTTTCAGATATTAATTTTC

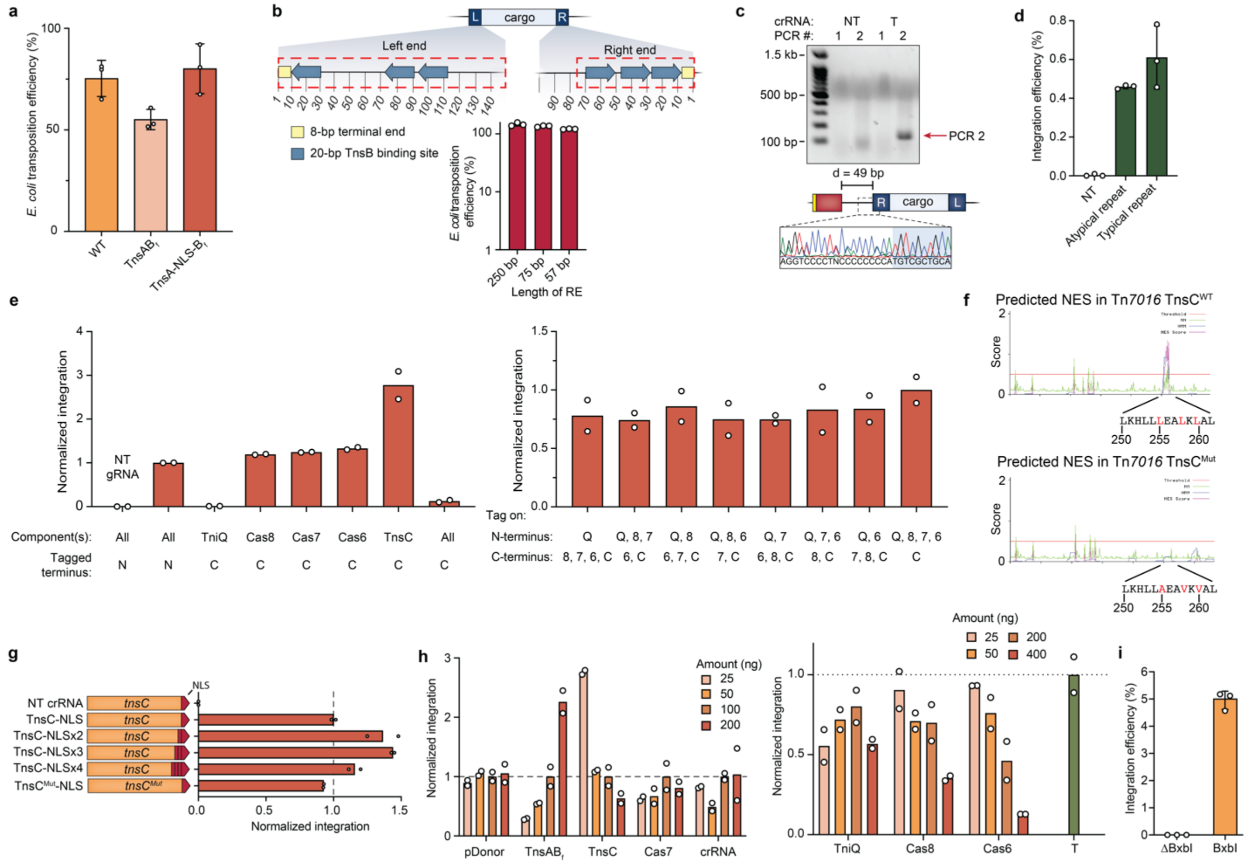
Supplementary Figure 4 | Detection of TnsC recruitment to a genomic locus and profiling of off-target binding events. a, 500 kb viewing window of ChIP-seq signal at the *TTN* promoter targeted by *TTN* Guide 1. **b,** Top, 5 kb viewing window of ChIP-seq peak at the *TTN* promoter targeted by *TTN* Guide 1. Bottom, 150 bp viewing window ChIP-seq peak at the *TTN* promoter targeted by *TTN* Guide 1. The peak summits in the targeting conditions align with the *TTN* promoter protospacer. **c,** Venn diagram showing overlap of targeting and non-targeting peaks. **d,** Heatmap of signal intensity in a 2 kb window surrounding the peak center in *TTN* targeting exclusive peaks (1203), sorted in descending order by mean signal over the window. The peak with the highest mean signal was at the *TTN* promoter, which was targeted by *TTN* Guide 1. **e,** Heatmap of signal intensity in a 2 kb window surrounding the peak center in non-targeting (NT) exclusive peaks (2526), sorted in descending order by mean signal over the window. ChIP-seq signal was weak across NT exclusive peaks. **f,** List of 5 genomic loci most similar to the *TTN* protospacer. Mismatches at every 6th nucleotide, denoted by an “X”, were disregarded due to the nature with which Cas7 binds to crRNAs. All other mismatches are shown in red. **g,** Manual inspection of a 10 kb window surrounding each predicted off-target sequence. Minimal enrichment of ChIP-seq signal was seen in either the *TTN* targeting or the non-targeting condition. Viewing windows in **a**, **b**, and **g** are shown for 3 biologically independent targeting and non-targeting samples, and ChIP-seq signal is visualized as signal per million reads (SPMR). Triangles in **a** and **g** denote the position of either the expected *TTN* targeting sequence or of the predicted mismatch sequences.



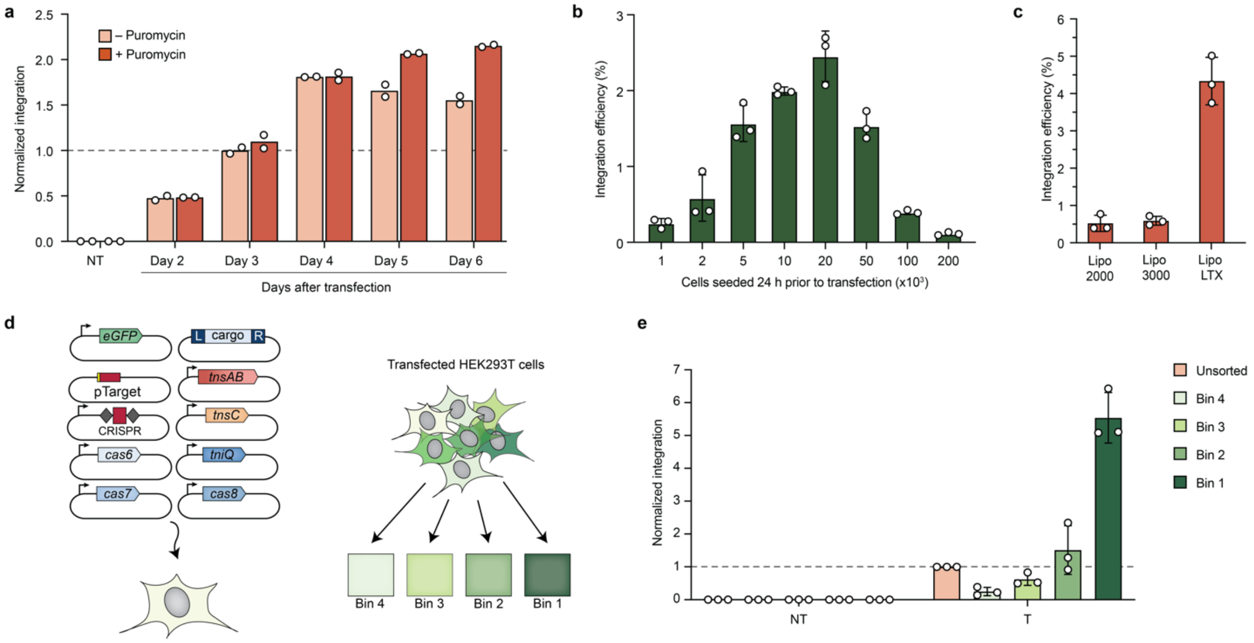
Supplementary Figure 5 | Initial detection and optimization of targeted integration using *Vch*CAST (eCAST-1). a, Initial quantification of ChlorR resistant *E. coli* colonies after isolation from human cells. **b,** Representative colony PCR of clonal integration products, detecting right transposon end (TnR) and left transposon end (TnL) junctions, as well as the KanR marker on the backbone of pTarget. Sanger sequencing of integration junctions are shown in **Figure 4b**. This was repeated in biological duplicate with similar results. **c,** Nested PCR strategy to detect plasmid-transposon junctions directly from HEK293T cell lysates (left), and agarose gel electrophoresis showing target-cargo junction product bands (right). Expected amplicon sizes are marked for each PCR reaction with red arrows, and the crRNA was either non-targeting (NT) or targeting (T). "H_2O" denotes a condition in which the lysate was omitted from the PCR reactions. An aliquot of PCR-1 is used for PCR-2 such that a "nested PCR" is performed (see Methods). Sanger sequencing was performed on the product after PCR-2 in the targeting condition (bottom right). This was repeated in biological triplicate with similar results. **d,** Schematic of TaqMan probe strategy used to improve signal-to-noise by selectively detecting novel plasmid-transposon junctions. Probes labeled with FAM (blue) are used to detect target-transposon junctions, and probes labeled with SUN (green) are used to detect the target plasmid backbone (reference sequence), for integration efficiency quantification. Probes that span the junction of pTarget and the right transposon end of eCAST-1 are designed to anneal to an insertion event 49-bp downstream of the target site. **e,** Integration efficiencies were improved by varying the relative levels of pDonor, pTarget, or protein expression plasmids, as indicated; data were normalized to a control sample transfected with 100 ng of each component (left), or to the 100 ng condition for each varied protein (right), which had an average value of either 0.004 % (left) or ranged from 0.0002–0.0005 % (right), respectively. Data in **e** are shown as mean for n = 2 biologically independent samples. Data shown in **e** were quantified via qPCR.



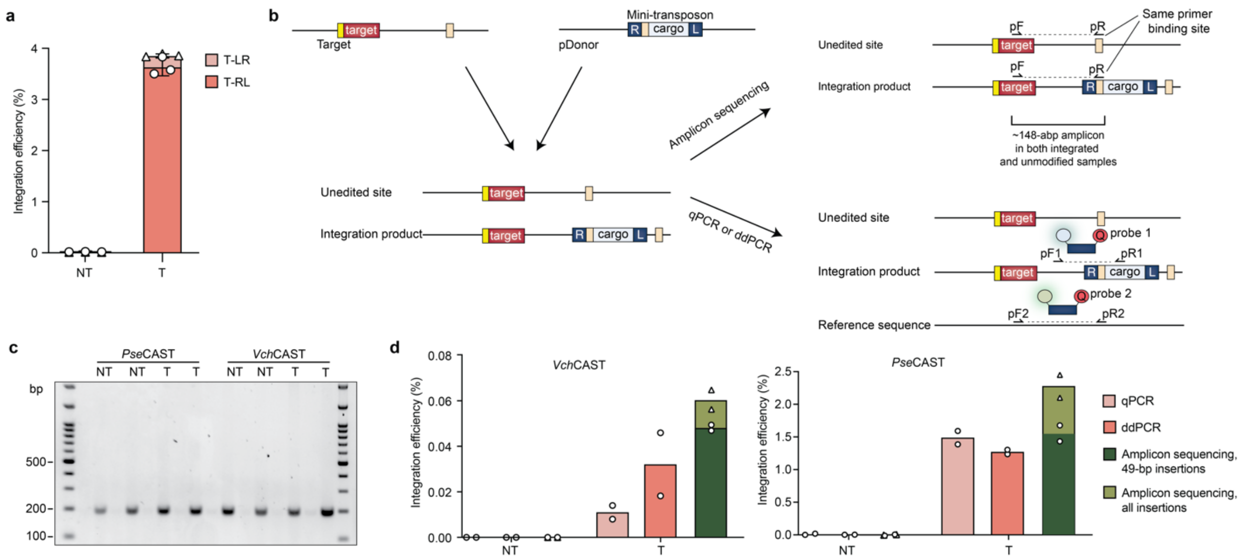
Supplementary Figure 6 | Systematic screening of homologous Type I-F CRISPR-associated transposons to uncover improved systems for mammalian cell applications. a, Cartoon depicting the multi-tiered approach that was applied to screen the indicated systems through a series of consecutive activity assays, with associated schematics shown for each functional assay. The middle panel depicts a transcriptional activation assay designed to monitor transposon DNA binding by TnsB in human cells using a tdTomato reporter plasmid. **b,** Western blotting to detect expression of candidate Cas6 homologs in HEK293T cells, with or without human codon optimization (hCO), using monoclonal anti-FLAG M2 antibody; β-actin was used as a loading control. We observed a range of expression levels for human codon-optimized gene variants, and genes were poorly expressed for most systems when native bacterial coding sequences were used. **c,** Activity assays for Cas6 homologs using the GFP knockdown assay shown in **Figure 1d**. For each homolog, GFP fluorescence levels were measured by flow cytometry and normalized to the experimental condition in which the GFP reporter plasmid lacked a CRISPR direct repeat (DR) in the 5'-UTR. **d,** Transcriptional activation data for TnsB-VP64 constructs from selected homologous CAST systems, as measured by flow cytometry. **e,** Transcriptional activation data for QCascade and TnsC-VP64 from homologous CAST systems, as measured by flow cytometry. Tn*7016*, the final homolog that was selected for additional screening for transposition, is marked with a red arrow. Data in **c–e** are shown as mean for n = 2 biologically independent samples. Uncropped western blots are shown in **Supplementary Figure 14.**



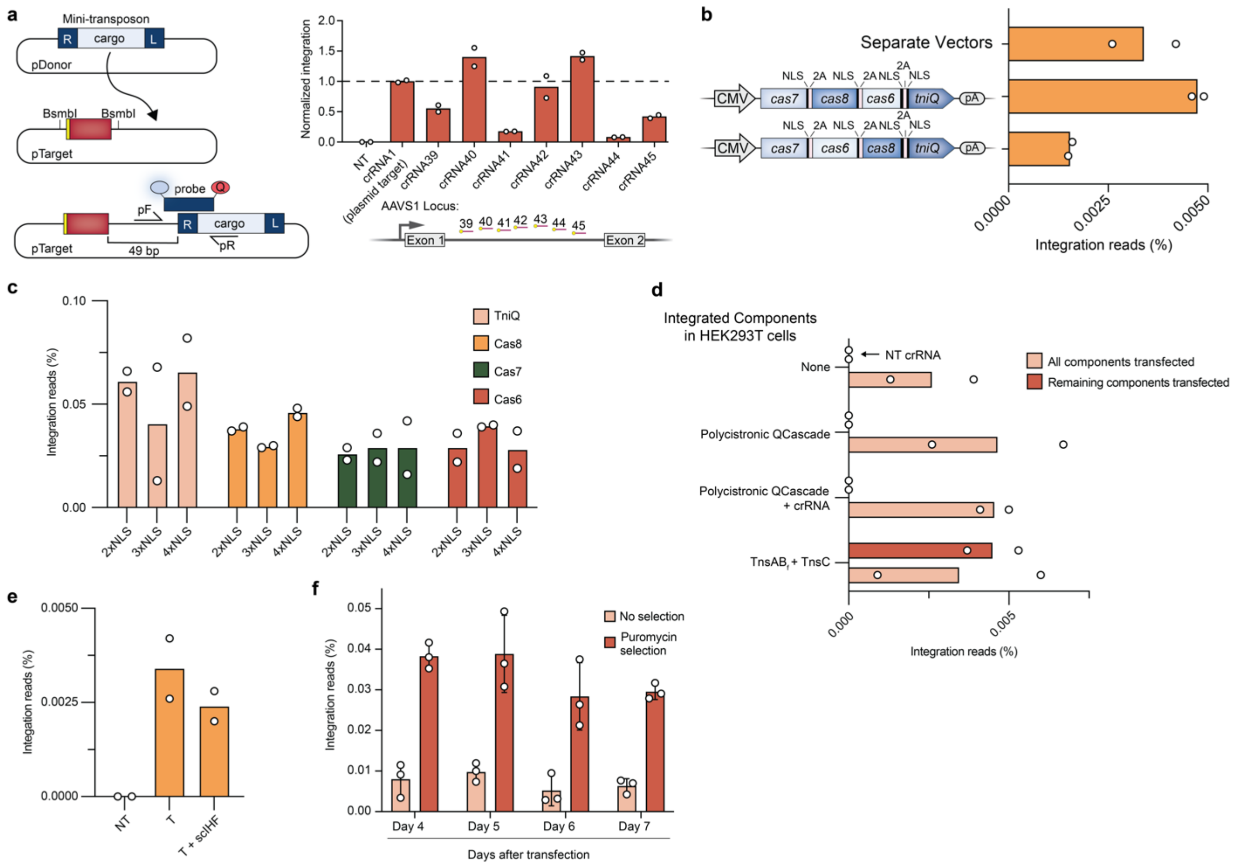
Supplementary Figure 7 | Parameter screening to further improve integration activity with the eCAST-2 (*Pse*CAST) system. a, RNA-guided DNA integration efficiency for TnsAB fusion (TnsAB$_f$) protein design, with or without internal NLS, compared to the wild-type TnsA and TnsB proteins. Experiments were performed in *E. coli*, and efficiencies were measured by qPCR. **b,** Tn*7016* transposon ends were shortened relative to the constructs tested previously[4], generating the constructs indicated with red dashed boxes at the top. RNA-guided DNA integration activity was compared for the indicated transposon right end (RE) variants in *E. coli*, as measured by qPCR (bottom), while a 145bp LE was used. The final pDonor design used in **Figure 4** contains 145-bp and 75-bp derived from the native left and right ends of *Pseudoalteromonas* Tn*7016*, respectively. **c,** Agarose gel electrophoresis showing successful junction products from nested PCR (top) for eCAST-2, and Sanger sequencing chromatograms showing the expected integration distance (bottom). **d,** Integration efficiencies in HEK293T cells were similar using either typical or atypical CRISPR repeats[5,6], as measured by qPCR. **e,** RNA-guided DNA integration activity was compared with the indicated BP NLS tags on eCAST-2 components, as measured by qPCR. Individual components had their respective BP NLS tag repositioned from the N- to the C-terminus; "All" represents a condition in which all components had BP NLS tags on the noted terminus (left). Interestingly, the observed tag sensitivity is similar to, but distinct from, that with eCAST-1 components. Various combinations of N- and C-terminal NLS tagging for *Pse*QCascade and *Pse*TnsC (right). NT = non-targeting crRNA. **f,** Nuclear export signal (NES) predictions for eCAST-2 wild type (WT) and mutant TnsC. We hypothesized that a putative NES within TnsC could led to inefficient nuclear localization, and selected multiple residues that, when mutated, might lower this risk. Predicted NES sequences were generated using NetNES. **g,** RNA-guided DNA integration activity was compared after appending additional NLS tags on *Pse*TnsC and removing a potential internal nuclear export signal (NES) sequence with the mutations L255A, L258V, and L260V, as indicated in **f**. **h,** RNA-guided DNA integration activity was compared after varying the relative levels of individual eCAST-2 protein and RNA expression plasmids. Data were measured by qPCR and are normalized to either the sample transfected with 100 ng of each component for each condition, with an average integration efficiency of 0.10–0.17% (left), or a control sample (labeled T) transfected with the standard eCAST-2 plasmid amounts, as detailed in the Methods section with an average integration efficiency of 2.7 % (right). **i,** A plasmid-based BxbI recombination assay was performed to benchmark eCAST-2 integration efficiency to other commonly used large DNA insertion tools. Data in **a, b**, **d,** and **i** are shown as the mean ± s.d. for n = 3 biologically independent samples. Data in **e**, **g**, and **h** are shown as the mean for n = 2 biologically independent samples.



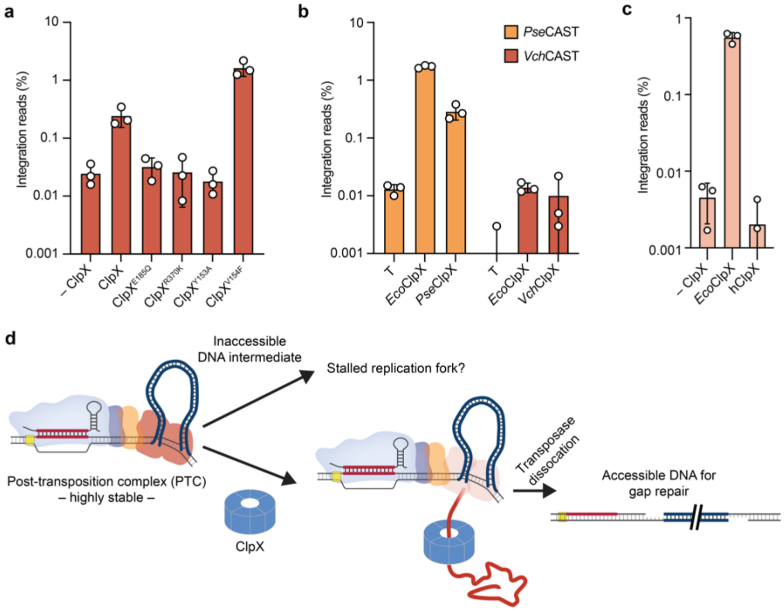
Supplementary Figure 8 | Selection, seeding, and sorting strategies result in further increases in eCAST-2.2 integration efficiencies. a, Normalized RNA-guided DNA integration efficiency for eCAST-2.2 in the absence or presence of puromycin selection, and after harvesting cells from between 2–6 days post-transfection. Experiments used a puromycin resistance plasmid as a transfection selection marker, in addition to eCAST-2.2 component plasmids, and integration activity was measured by qPCR and normalized to the condition harvested on day 3 without puromycin selection, which had an average integration efficiency of 2.3 %. **b**, eCAST-2.2 integration efficiencies were compared as a function of seeding density 24 hours before transfection. 24-well plates were with various cell densities ranging from 1 x 10^3 to 2 x 10^5 cells per well, and integration activity was measured by qPCR. **c**, Transfection of HEK293T cells via various cationic lipid delivery methods affected integration efficiencies. **d**, Schematic showing the use of a GFP transfection marker and cell sorting to increase integration efficiency. A GFP expression plasmid was transfected in significantly smaller amounts relative to eCAST-2.2 component plasmids, and cells were sorted into bins of varying GFP expression levels. **e**, eCAST-2.2 integration efficiencies are enhanced after using flow cytometry to sort cells for the brightest GFP positive cells. Cells were sorted four days after transfection, and the top 20% brightest cells were binned in increments of 5%, with Bin 1 representing the top 5% brightest cells and Bin 4 representing the 15-20% brightest cells. Integration efficiencies were determined for each bin separately, or for the unsorted population, as measured by qPCR. Integration efficiencies were normalized to the unsorted, targeting crRNA condition, which had a value of 0.44 %. Data in **a** are shown as the mean of n = 2 biologically independent samples. Data in **b**, **c**, and **e** are shown as the mean ± s.d. for n = 3 biologically independent samples.



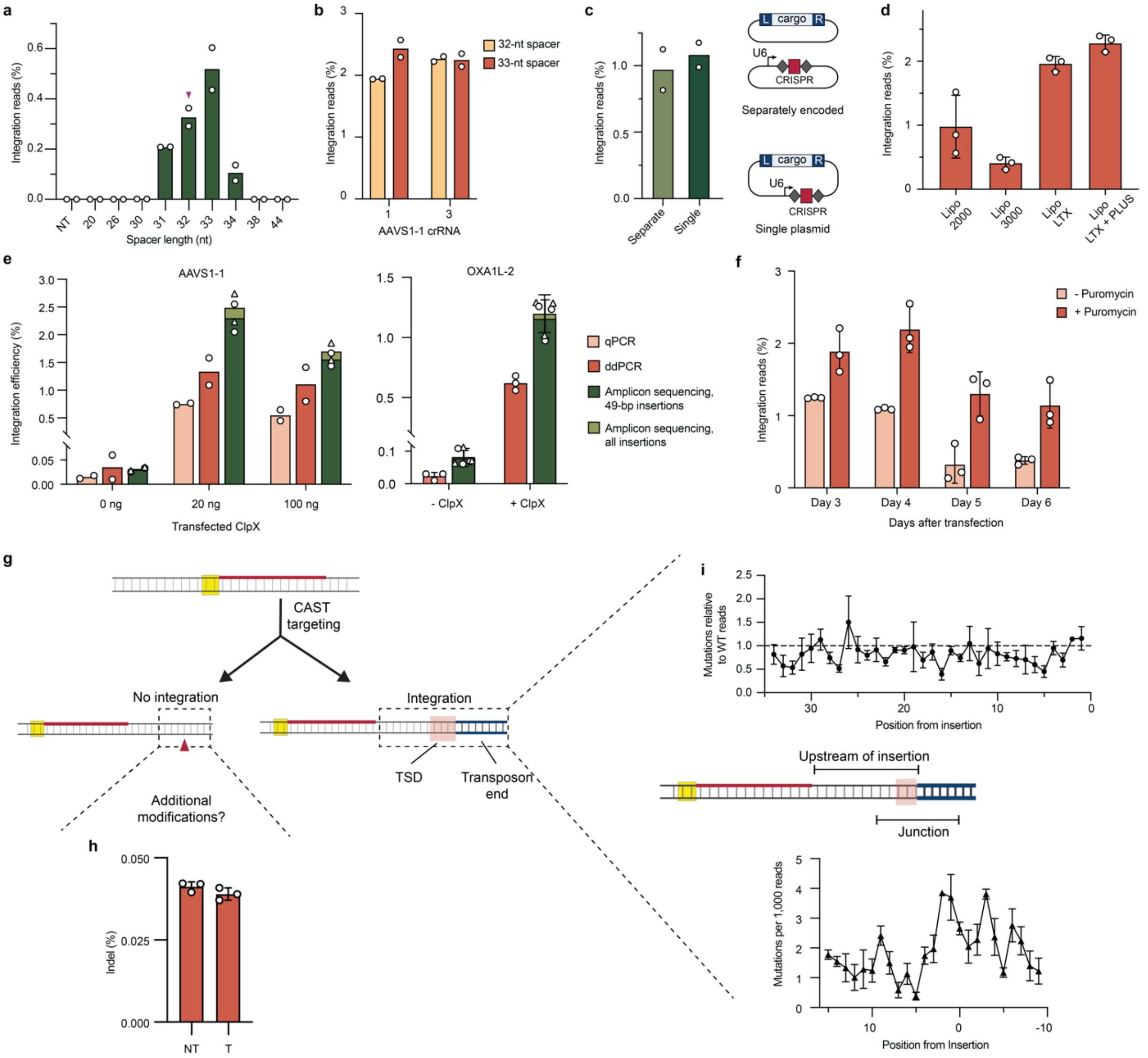
Supplementary Figure 9 | eCAST-2.2 integration is biased towards T-RL insertion and reproducibly quantified across distinct approaches. a, RNA-guided DNA integration is heavily biased towards insertion in the right-left (T-RL) orientation, with only a small minority of insertion events occurring in the left-right (T-LR) orientation. Integration efficiencies were calculated using SYBR qPCR. Triangle data points represent integration events in the T-LR orientation, while circle data points represent integration events in the T-RL orientation. **b**, Comparison of different strategies to detect and quantify integration efficiencies. For next-generation amplicon sequencing, a variant pDonor was constructed in which a primer binding site that is also present at the target site is cloned within the transposon cargo at a distance from the transposon right end (R), such that unedited sites and integration products yield amplicons of indistinguishable length using pF and pR primers (top). Consequently, next-generation sequencing of these amplicons provides relative abundances of edited and unedited alleles in the population, allowing for higher sensitivity in detecting integration efficiencies. For qPCR and ddPCR detection strategies, Taqman probes and primers are designed to amplify either the integration product or a reference sequence used to calculate integration efficiencies (bottom). For plasmid-based integration assays, the reference sequence is a distinct sequence on the plasmid target (pTarget). **c,** Representative agarose gel electrophoresis demonstrates identical amplicon products for non-targeting (NT) and targeting (T) samples after PCR-1 for NGS analysis. This was repeated in biological triplicates with similar results. **d**, Calculated integration efficiencies for the same experimental samples, measured by TaqMan qPCR, droplet digital PCR (ddPCR), and amplicon deep sequencing. ddPCR and qPCR analyses specifically probe for integration products that are 49-bp downstream of the target site, whereas amplicon sequencing analysis does not impose the same stringent distance bias, allowing the quantification of integration products within a larger window surrounding the anticipated integration site. Editing efficiencies for both eCAST-2.2 and eCAST-1 were consistent between different quantification methods. For amplicon sequencing samples, triangle data points represent all insertions characterized, while circle data points represent only 49-bp insertions. Data in **a**, are shown as the mean ± s.d. for n = 3 biologically independent samples. Data in **c** are shown as the mean for n = 2 biologically independent samples.



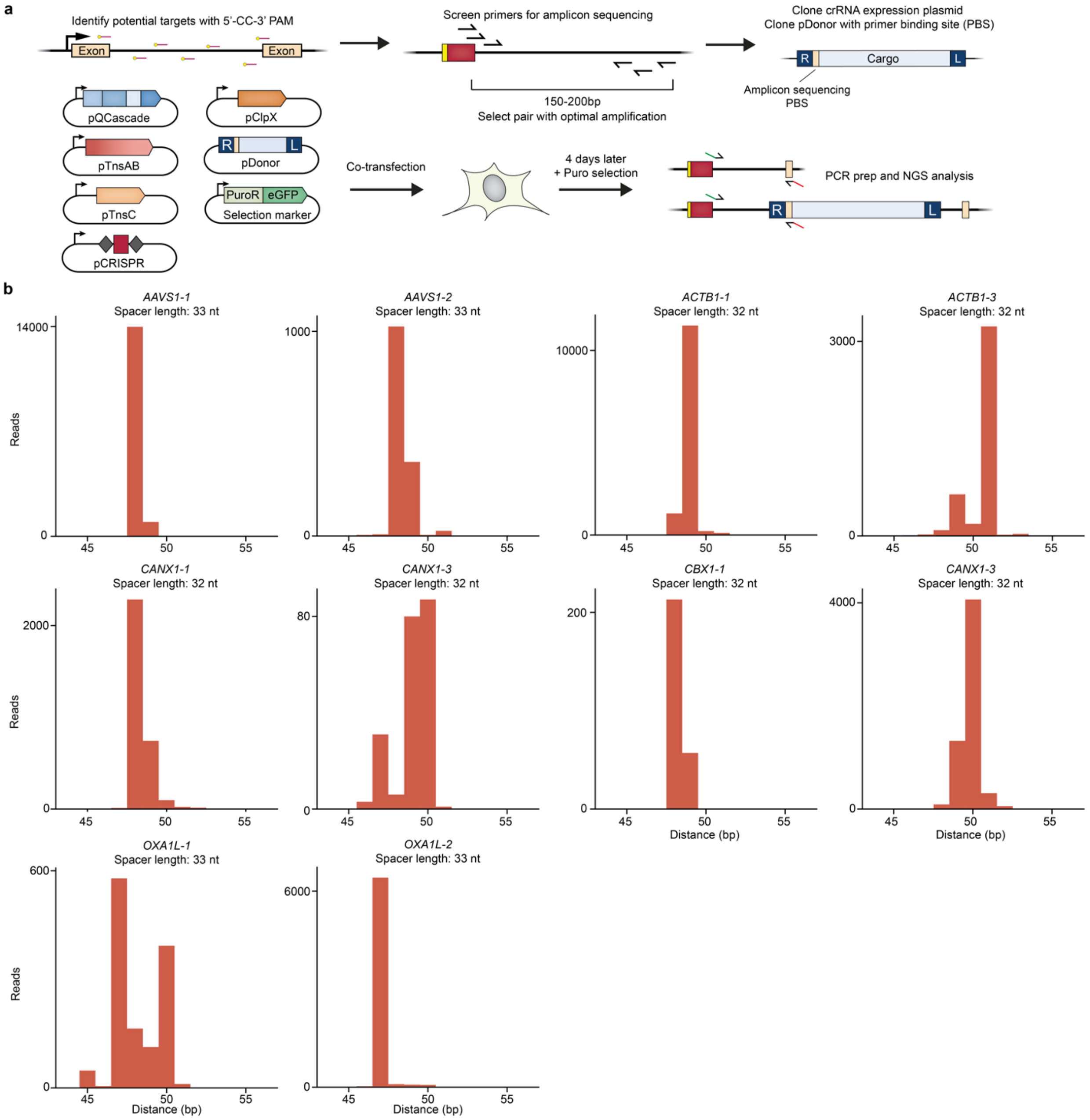
Supplementary Figure 10 | Preliminary attempts to improve eCAST-2.2 genomic integration activity and identify kinetic bottlenecks. a, A unique target site was cloned into a modified pTarget, in which the downstream integration site sequence remained the same, allowing us to specifically understand the impact of different crRNA sequences on integration efficiencies (left). Cloning various target sites into our modified pTarget that correspond to target sites within the AAVS1 safe harbor locus enabled screening of crRNAs to identify active sequences (right). Efficiencies were normalized to the crRNA used in our plasmid-targeting assays, which had an average integration efficiency of 2.0 %. **b**, Simplification of transfection workflow via polycistronic expression of QCascade, and genomic integration efficiencies with different constructs. "Separate Vectors" represents a condition in which *TniQ, Cas8, Cas7,* and *Cas6* were all expressed from separate pcDNA3.1-like vectors. **c,** Impact of additional NLS tags on eCAST-2 QCascade components on genomic integration efficiencies. All QCascade components had a singular NLS tag, unless noted. **d,** Impact of stably-expressed eCAST-2 components on genomic integration efficiencies. Cell lines were generated via Sleeping Beauty with drug selection, and various components were stably expressed (indicated by operons shown on the y-axis). "All components transfected" represents conditions in which all eCAST-2 components were co-transfected, while "Remaining components transfected" represents conditions in which only the non-expressed eCAST-2 components were transfected. **e,** Impact of co-transfection of *E. coli* Integration Host Factor (IHF) on human genomic integration efficiencies. "T + scIHF" represents a condition in which a plasmid expressing a single-chain IHFa/b[7] was co-transfected with a targeting gRNA. **f,** Varying cell harvest day and selection of transfected cells based on a concurrent drug marker improves integration efficiencies, although overall efficiencies remain low. Data in **a–e** are shown as mean for n = 2 biologically independent samples. Data in **f** are shown as the mean ± s.d. for n = 3 biologically independent samples. Data in **a** was determined by qPCR. Data in **b–f** were determined by amplicon sequencing.



Supplementary Figure 11 | Genomic editing outcomes with ClpX. a, Mutational analysis of ClpX-mediated editing improvements. Point mutations were designed to either ablate ATP hydrolysis (E185Q and R370K)[8,9] or perturb substrate engagement (Y153A and V154F)[10]; prior work demonstrated that Y153A ablates Mu replication *in vivo*, while V154F does not affect Mu growth, but specifically inhibits binding and degradation of ssrA-tagged substrates[10]. **b,** The impact of native ClpX proteins on eCAST-2 and eCAST-1. *Pse*ClpX and *Vch*ClpX improved eCAST-2 and eCAST-1 genomic integration efficiencies, respectively, but *Eco*ClpX consistently produces a more robust improvement. **c,** Human-derived ClpX does not improve genomic integration efficiencies for eCAST-2. The putative mitochondrial targeting sequence from human derived ClpX[11,12] was replaced with a BP-NLS tag. **d,** Proposed model for the role of ClpX in improving genomic integration efficiencies. In the absence of ClpX, the PTC is sufficiently stable to prevent accessibility to the DNA intermediate, leading to a loss of genomic integration events. In contrast, inclusion of ClpX facilitates unfolding of CAST components, resulting in destabilization/dissociation of the complex and accessibility to the DNA intermediate. Data in **a, b,** and **c** are shown as the mean ± s.d. for n = 3 biologically independent samples.



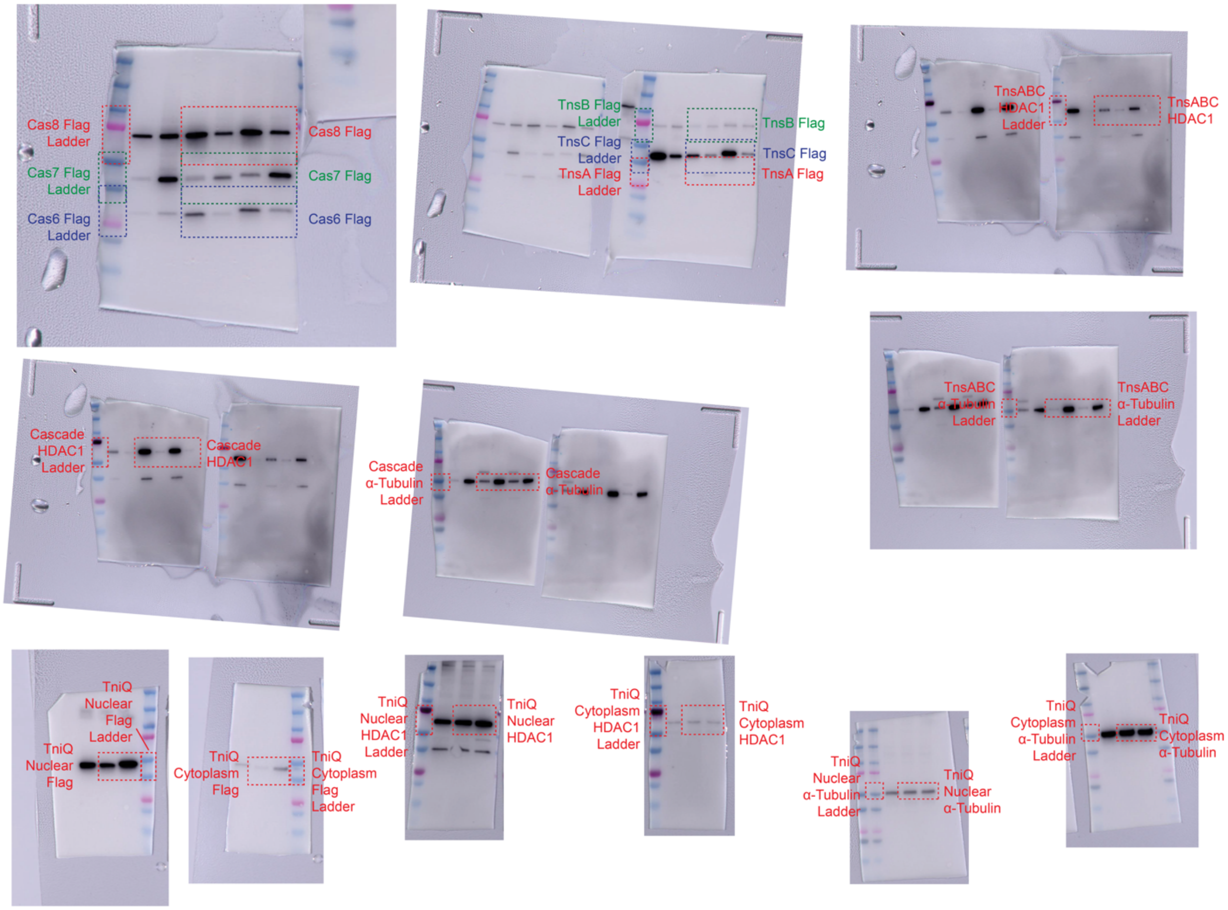
Supplementary Figure 12 | Engineering CAST systems with ClpX. a, Impact of atypical spacer lengths on plasmid-based integration efficiencies (the canonical spacer length, 32nt, is marked with a maroon triangle). **b,** Impact of 32nt vs 33nt spacer lengths on genomic integration efficiencies at the AAVS1-1 target site. Two different crRNAs were tested that were nearby in the genomic locus, minimizing disruption of potential downstream integration-site requirements. **c,** Impact of encoding the crRNA on the pDonor for genomic integration efficiencies. The U6 promoter, crRNA, and U6 terminator sequences were cloned on either a separate plasmid or in the pDonor backbone. **d,** Genomic integration as a function of different cationic lipid transfection methods **e,** A comparison of integration efficiencies in the presence and absence of ClpX as measured by qPCR, ddPCR, and amplicon sequencing for AAVS1-1; ddPCR and amplicon sequencing for OXA1L-2. For amplicon sequencing samples, triangle data points represent all insertions characterized, while circle data points represent only 49-bp insertions. **f,** Varying cell harvest day and selection of transfected cells based on a concurrent drug marker improves integration efficiencies, in the presence of ClpX **g,** Schematic of sequences that were analyzed to understand if undesirable editing outcomes were occurring with eCAST-3. If a sequence did not contain a transposon end, the sequence surrounding the intended integration site was investigated for a higher frequency of indel events compared to samples in which a non-targeting crRNA was used. If a transposon end was detected in the sequence, the sequence was analyzed for additional mutations. **h,** Mutations surrounding the integration region at AAVS1-1 do not occur above background frequencies present when a NT crRNA is co-transfected. **i,** Mutations upstream the integration site at AAVS1-1 do not occur at a higher rate compared to WT alleles (top). Mutations in the transposon end and surrounding the target site duplication at AAVS1-1 do not occur at rates above background sequencing error (bottom). Integration events at the major integration site (49bp downstream of crRNA) were analyzed. Data in **a-c** and **e** (for AAVS1-1) are shown as mean for n = 2 biologically independent samples. Data in **d, e** (for OXA1L-2), **f, h,** and **i** are shown as mean ± s.d. for n = 3 biologically independent samples. Data were quantified by amplicon sequencing.



Supplementary Figure 13 | Leveraging eCAST-3 to perform targeted RNA-guided DNA integration at multiple target sites. a, Current workflow for applying eCAST-3 to new target sites. First, potential targets with CC PAMs are identified in region of interest. Target sites are then screened for optimal primers for amplicon sequencing. The downstream primer binding site is cloned into a pDonor immediately adjacent to the RE, enabling NGS-based quantification. Cells are then transfected with pCRISPR, pQCascade, pTnsAB, pTnsC, pClpX, pDonor, and an optional drug selection marker. After 4 days, cells can be harvested for PCR prep and subsequent NGS-based analysis. **b,** Representative integration site distributions for transfections shown in **Figure**

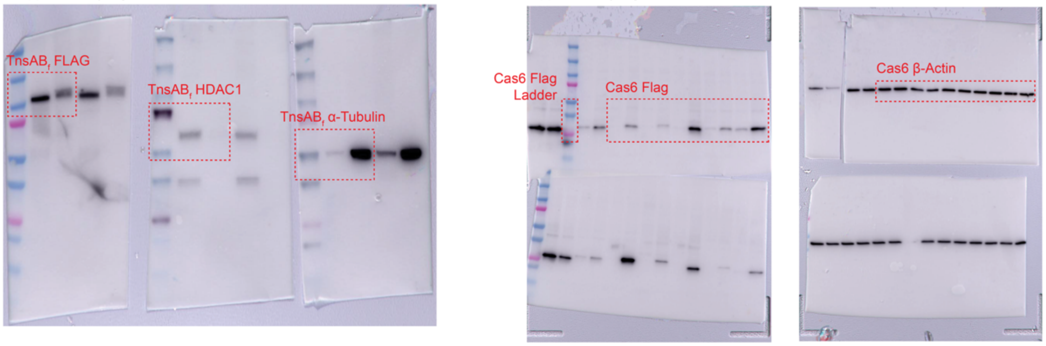
5i. The length of the spacer is shown, and the distance represents the length from the PAM-distal end of the spacer to the transposon end.

a Supplementary Figure 1a

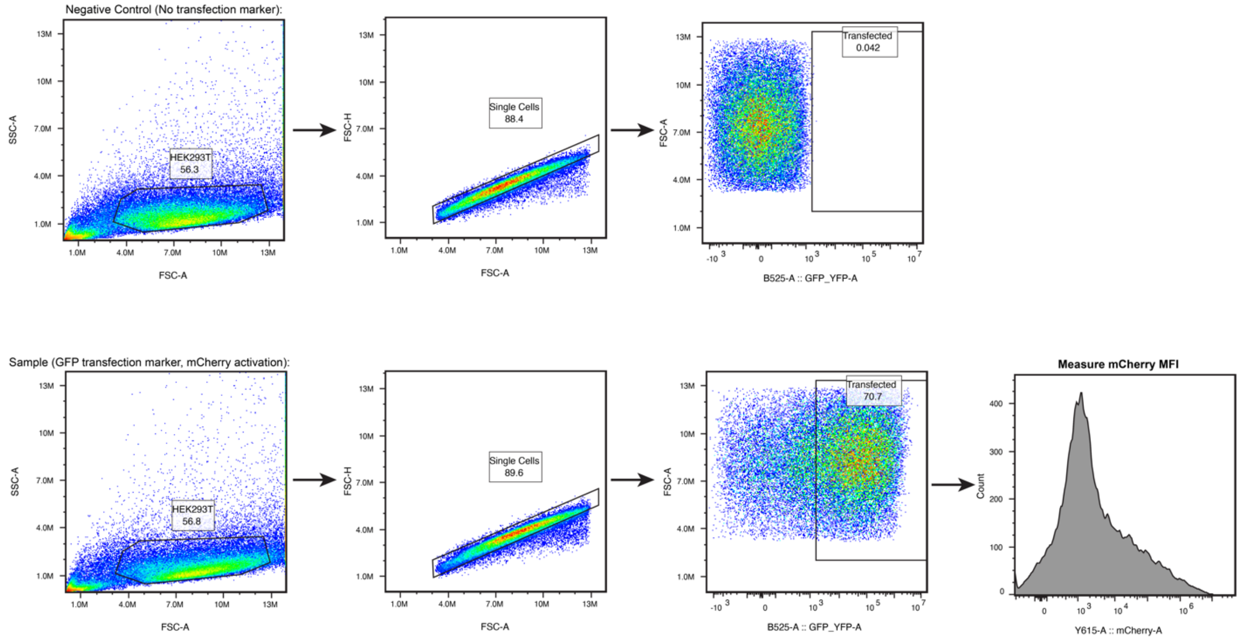


b Supplementary Figure 1d

c Supplementary Figure 6a



Supplementary Figure 14 | Uncropped Western blot images. a, Uncropped membranes for **Supplementary Figure 1a. b,** Uncropped membranes for **Supplementary Figure 1d. c,** Uncropped membranes for **Supplementary Figure 6a**.



Supplementary Figure 15 | Representative analysis of flow cytometry samples. a, Cells are first gated for cells versus debris, then for single cells, and lastly for transfected cells. A negative control population is used to accurately define cells that are positive for the transfection marker. Finally, the mean fluorescence intensity (MFI) is measured for experimental samples in the desired fluorescent channel.

SUPPLEMENTARY REFERENCES

1. Cameron, P. *et al.* Harnessing type I CRISPR–Cas systems for genome engineering in human cells. *Nat. Biotechnol.* **37**, 1471–1477 (2019).
2. Perez-pinera, P. *et al.* RNA-guided gene activation by CRISPR-Cas9 – based transcription factors. *Nat. Methods* **10**, 973–976 (2013).
3. Chavez, A. *et al.* Highly-efficient Cas9-mediated transcriptional programming. *Nat. Methods* **12**, 326–3228 (2015).
4. Klompe, S. E., Vo, P. L. H., Halpin-Healy, T. S. & Sternberg, S. H. Transposon-encoded CRISPR–Cas systems direct RNA-guided DNA integration. *Nature* **571**, 219–225 (2019).
5. Klompe, S. E. *et al.* Evolutionary and mechanistic diversity of Type I-F CRISPR-associated transposons. *Mol. Cell* **82**, 616-628.e5 (2022).
6. Petassi, M. T., Hsieh, S. & Peters, J. E. Guide RNA Categorization Enables Target Site Choice in Tn7-CRISPR-Cas Transposons. *Cell* **183**, 1757-1771.e18 (2020).
7. Corona, T. *et al.* Activation of site-specific DNA integration in human cells by a single chain integration host factor. *Nucleic Acids Res.* **31**, 5140–5148 (2003).
8. Joshi, S. A., Hersch, G. L., Baker, T. A. & Sauer, R. T. Communication between ClpX and ClpP during substrate processing and degradation. *Nat. Struct. Mol. Biol.* **11**, 404–411 (2004).
9. Hersch, G. L., Burton, R. E., Bolon, D. N., Baker, T. A. & Sauer, R. T. Asymmetric interactions of ATP with the AAA+ ClpX6 unfoldase: Allosteric control of a protein machine. *Cell* **121**, 1017–1027 (2005).
10. Siddiqui, S. M., Sauer, R. T. & Baker, T. A. Role of the processing pore of the ClpX AAA+ ATPase in the recognition and engagement of specific protein substrates. *Genes Dev.* **18**, 369–374 (2004).
11. Santagata, S. *et al.* Molecular cloning and characterization of a mouse homolog of bacterial ClpX, a novel mammalian class II member of the Hsp100/Clp chaperone family. *J. Biol. Chem.* **274**, 16311–16319 (1999).
12. Kang, S. G. *et al.* Functional proteolytic complexes of the human mitochondrial ATP-dependent protease, hClpXP. *J. Biol. Chem.* **277**, 21095–21102 (2002).