Title

RNA-guided genome engineering: paradigm shift towards transposons

Authors

Chang, Chin-Wei
Truong, Vy Anh
Pham, Nam Ngoc
et al.

Publication Date

2024-03-01

DOI

10.1016/j.tibtech.2024.02.006

RNA-guided Genome Engineering: Paradigm Shift Towards Transposons

Chin-Wei Chang[1, ξ], Vy Anh Truong[1, ξ], Nam Ngoc Pham[1], and Yu-Chen Hu[1,2*, ORCID]

[1] Department of Chemical Engineering, National Tsing Hua University, Hsinchu, Taiwan 300
[2] Frontier Research Center on Fundamental and Applied Sciences of Matters, National Tsing Hua University, Hsinchu, Taiwan 300

Keywords: CASTs, OMEGA, transposon, RNA-guided genome engineering, CRISPR, double strand break

Running Title: RNA-guided Transposons for Genome Editing

[ξ] These authors contributed equally to this work.
*Correspondence: ychu@mx.nthu.edu.tw (Y.-C. Hu)
ORCID https://orcid.org/0000-0002-9997-4467

Abstract

CRISPR-Cas systems revolutionize the genome engineering field but need to induce double strand break (DSB) and may be difficult to deliver due to the large protein size. In 2019, Tn7-like transposons such as CASTs (CRISPR-associated transposons) are repurposed for RNA-guided DSB-free integration. In 2021, OMEGA (obligate mobile element guided activity) proteins of the IS200/IS605 transposons are developed as the hypercompact RNA-guided genome editing tools. The characteristics of these systems generate excitement for leveraging CASTs and OMEGA as innovative genome engineering tools and exploring methods to improve the precision and efficiency of editing. This review explores the recent developments and uses of CASTs and OMEGA in genome editing across prokaryotic and eukaryotic cells. The pros and cons of these transposon-based systems are deliberated in comparison to other CRISPR systems.

CRISPR and Transposons for Genome Engineering

CRISPR-Cas systems are categorized into 2 classes, 6 types (e.g. Cas9 for type II and Cas12 for type V) and 33 subtypes [1], among which CRISPR-Cas9 is first repurposed for genome engineering of eukaryotes [2, 3] and prokaryotes [4, 5]. CRISPR-Cas9 requires **Cas9** (See Glossary), **crRNA** and **tracrRNA** (See Glossary) to recognize the **PAM** (See Glossary) sequence on the target DNA and induce double strand break (DSB), which triggers intrinsic **DNA repair** (See Glossary) and ensuing DNA insertions/deletions (indels). However, DNA repair is inefficient in non-dividing eukaryotes [6] and some bacteria [7]; DSB sparks innate immune responses in human cells [8] and may elicit genomic deletions and chromosomal translocations [9, 10]. Furthermore, the large size of the most widely used Cas9 and Cas12 (1000-1300 aa) hinders their delivery by **AAV** (See Glossary) for gene therapy [11, 12]. Although miniature Cas12f homologs (400–700 aa) are found, their editing efficiencies remain low [13-15].

In 2017, the connection between Tn7-like **transposons** (See Glossary) and CRISPR is discovered [16], with subsequent observations revealing associations with various CRISPR types, including I-B, I-D, I-F, and V-K [17-22]. In 2019, the type V-K [17] and type I-F3 [18] CRISPR-associated transposons (termed CASTs) are first repurposed for DSB-free, RNA-guided DNA insertion into *E. coli*, which implicates the potentials of CASTs to obviate the requirement of DSB induction in CRISPR-Cas systems. In 2021, IscB (400-500 aa) and TnpB (408 aa) nucleases in the **insertion sequence (IS)** (See Glossary) family transposons are transformed into hypercompact programmable RNA-guided DNA nucleases [23, 24]. These nucleases, part of the OMEGA (obligate mobile element guided activity) system, show promise in overcoming the size limitations of current CRISPR-Cas nucleases.

Several articles have comprehensively reviewed CASTs, either offering a general introduction [25-28] or delving into specific aspects like molecular mechanisms [20, 21, 29] and applications related to DSB-free DNA integration [28, 30, 31]. Yet OMEGA system is reviewed in only two papers [21, 32] and several research spotlight articles [33-35]. This paper provides a comprehensive review of recent advances, specifically focusing on the development and applications of CASTs in genome editing for both prokaryotes and eukaryotes. Additionally, we offer an updated overview of OMEGA in the context of eukaryotic genome editing.

What are CASTs and OMEGA

Bioinformatic searches have revealed CASTs repertoire to over 1000 subsystems [36], among which types V-K and I-F are the most exhaustively studied [26]. Type V-K CASTs encode four proteins (TnsB, TnsC, TniQ, Cas12k), tracrRNA/crRNA and the genetic cargo within the left end (LE) and right end (RE) of transposon (Fig. 1A). TnsB is the transposase for integration; TnsC is an ATPase for target site selection while TniQ is an adaptor protein. Type I-F3 CASTs encode seven proteins (TnsA, TnsB, TnsC, Cas6, Cas7, Cas5/8 and TniQ) and expresses only crRNA (without tracrRNA). The cassette, along with the cargo, is embedded within the LE/RE ends (Fig. 1A).

OMEGA is a new family of miniature RNA-guided endonucleases comprising TnpB, IscB, IsrB and Fanzor [32]. IscB and TnpB prevent permanent transposon loss [37] and are evolutionary ancestors of Cas9 and Cas12 [38], respectively (Fig. 1B). IscB shares HNH and RuvC nuclease domains with Cas9 [23] and is guided by ωRNA, enabling IscB to cleave its target DNA (Fig. 1B). Similarly, TnpB associates with ωRNA for dsDNA cleavage (Fig. 1B) [24]. Recently, Fanzors found in eukaryotic transposons are utilized as the first eukaryotic OMEGA RNA-guided DNA endonucleases for programmable human genome engineering [39].

CASTs for Prokaryotic Genome Engineering

Development of type V-K CASTs

Strecker, *et al.* first characterize the type V-K CAST from *Scytonema hofmanni* (*Sh*CAST) and design **gRNAs** (see Glossary) for DNA insertion into *E. coli* [17]. The insertions occur unidirectionally with the requirement of a targeting protospacer and a PAM sequence. The authors construct a helper plasmid (pHelper) expressing all 4 *Sh*CAST proteins and a donor plasmid (pDonor) carrying the genetic cargo flanked by LE/RE (Fig. 2A). Co-electroporation of both plasmids into *E. coli* confers integration with varying efficiencies (Table 1). *Sh*CAST requires Cas12k to coordinate with gRNA and recruit the transposase complex to integrate DNA cargo (up to 10 kb) into the target site ≈60–66 bp downstream of the PAM, without triggering DSB [17].

Note that *Sh*CAST induces the integration of LE/RE and is not a scarless method. Although the PAM and target sequence remain unchanged post-transposition, they prevent additional integration at the same site [17], resembling "target immunity". Under overexpression conditions, *Sh*CAST results in off-target integration through Cas12k-independent mechanisms [17, 28]. Moreover, *Sh*CAST mediates co-integration of not only the transposon but also the donor plasmid [17, 40-42]. Such co-integration is observed in multiple *E. coli* strains [42].

*Sh*CAST is subsequently improved by using different promoters in the pHelper plasmid (Table 1), which shows that stronger promoter and higher expression improve integration efficiency [43]. By designing the crRNA to target elements that naturally exist in multiple copies in bacteria, *Sh*CAST enables multiplexed gene insertion into 4 identical repeats and integration of a large DNA cargo into genomic sites (Table 1).

Increasing Cas12k expression level alone enhances on-target integration efficiency [42, 44]. Accordingly, the SHOT system (Fig. 2A, Table 1) is developed by increasing Cas12k expression

and tuning editing conditions [42]. SHOT can insert genes into loci that cannot be edited using the original *Sh*CAST [42]. Other *Sh*CAST-based systems such as C12KGET [45] is also developed by modifying the expression cassette designs (Table 1).

These studies confirm that *Sh*CAST can be adapted and improved for DSB-free, multiplexed integration of large DNA cargo into prokaryotic genome. Note that the insertion efficiencies vary widely with the genomic sites, probably because the sequences directly adjacent to the LE/RE are pivotal for the integration efficiency [46]. The stoichiometry of individual proteins is also crucial for editing efficiencies, which can be explained by recent discoveries of structures and transposition mechanisms (Box 1 and Fig. 2B). Tuning the expression levels of each component proteins can enhance the editing efficiencies and reduce the off-target effects, yet co-integration of plasmid backbone remains problematic.

This co-integration problem is addressed in a recent study [46]. Bacterial transposons propagate through cut-and-paste or copy-and-paste pathway [47]. In the Tn7 transposon, TnsA and TnsB nick the 5' and 3' end of donor DNA, respectively, resulting in simple insertions via cut-and-paste transposition. Tou *et al*. reason that the lack of TnsA in the type V-K CAST leads to TnsB-mediated 3' nicking only, thus resulting in copy-and-paste integration and undesired co-integration [46]. With this hypothesis, HELIX (Table 1) is developed from *Sh*CAST by fusing a nicking endonuclease (nAnil) to the N terminus of TnsB to restore the 5' nicking capability (Fig. 2A). HELIX confers large DNA insertion into genomic targets in *E. coli* via cut-and-paste mechanism, which almost abolishes the co-integration while retaining robust integration efficiencies (Table 1). Note, however, that C-terminal fusion of nAnil to TnsB does not yield transposition. The HELIX design, derived from ShCAST, is also applicable to other type V-K CASTs or a different strain of *S. hofmannii* [46].

Altogether, HELIX streamlines the type V-K transposition and enables programmable, cut-and-paste, unidirectional and recombination-independent DNA insertions. Note, however, that engineering of the component proteins is sensitive to perturbations.

Development of Type I CASTs

At the same time, Klompe *et al*. [18] demonstrate programmable transposition in *E. coli* using the type I-F3 CAST from *Vibrio cholera* (*Vch*CAST) and establishes the INTEGRATE system. INTEGRATE encode all required proteins and donor DNA in 3 separate plasmids: pQCascade, pTnsABC and pDonor (Fig. 3A and Table 1). INTEGRATE enables large DNA integration into *E. coli* genome in either forward or reverse direction with a 5-bp target site duplication (TSD), and accurate DNA integration into *E. coli* genome 46–55 bp downstream of the PAM-distal end target site (Table 1 and Fig. 3B). Similar INTEGRATE systems based on type I-B CASTs are also described [48, 49].

Following the concept of INTEGRATE, MUCICAT (Table 1) systems based on different type I-F CASTs is developed for multicopy and multiplexed integration by changing the crRNA design [50, 51]. Yet, the multicopy integration does not function as effectively in *E. coli* MG1655ΔendAΔrecA(DE3) [50]. Another type I-F CAST also fails to mediate cargo insertion [51]. These studies indicate the need to screen appropriate type I CASTs from different organisms for INTEGRATE system development.

To streamline INTEGRATE, Vo, *et al*. assemble all components of INTEGRATE in an all-in-one expression vector (Fig. 3A) [52]. The all-in-one INTEGRATE confers higher integration efficiencies than the original INTEGRATE, presumably because the higher efficiency of transforming a single plasmid than co-transforming 3 plasmids into single cells. DNA insertion is directional and off-target effect is lower than that of the original version. The all-in-one type I-F3

158 INTERGRATE is further harnessed with an orthogonal all-in-one type V-K *Sh*CAST [52] to
159 facilitate iterative insertions and circumvents the target immunity problem of *Sh*CAST. Using a
160 gRNA array, the all-in-one INETGRATE achieves multiplexed insertions into *E. coli* genomic loci
161 and can be used in other bacteria [52]. Similarly, all-in-one MUCICAT is generated and operated
162 orthogonally with all-in-one *Vch*CAST for multiplexed gene insertion [51]. Rubin, *et al*. further
163 develop a DNA-editing All-in-one RNA-guided CRISPR–Cas Transposase from *V. cholerae*
164 (*Vc*DART) [53]. *Vc*DART resembles the all-in-one INTEGRATE but the genetic cargo is barcoded
165 for environmental transformation sequencing (ET-seq), which enables tracking of edited cells.
166 These studies highlight the broad utility of all-in-one type I CASTs for bacterial engineering.

167 Overall, type I-F CASTs confer higher integration specificity (i.e. no co-integration) than type
168 V-K CASTs due to their differences in transpososome structures and transposition mechanisms
169 (Box 1 and 2). Both CASTs can be cloned into a single plasmid for enhanced efficiency, specificity,
170 and orthogonal editing, indicating a growing preference for all-in-one CASTs. Although type V-K
171 CASTs yield more serious off-target integration and co-integration, these issues can be addressed
172 by the HELIX design, supplementation of additional components such as pir [46] and fine-tuning
173 the stoichiometry of component proteins. The intrinsic target immunity problems of type V-K
174 CASTs can be circumvented by using orthogonal CASTs. Type V-K CASTs require only 4 proteins
175 and is more compact than type I-F CASTs that necessitate 7 proteins, which makes it easier to
176 construct and deliver the all-in-one vector. Consequently, it is anticipated that type V-K CASTs
177 will become more popular for prokaryotic genome editing compared to type I-F.

Applications of CASTs in Bacteria

Metabolic engineering for improved product synthesis

E. coli BL21(DE3) is commonly used for protein production, but excess acids byproduct accumulation is a problem. Chang, *et al*. design a 10.3 kb CRISPR interference (CRISPRi) module that concurrently suppresses 4 genes accountable for acid accumulation [42]. Integration of the CRISPRi module into *E. coli* BL21(DE3) using SHOT (Table 1) represses all 4 genes and acid accumulation, hence producing more recombinant protein [42], and making this strain a promising platform for recombinant protein production. The C12KGET system (Table 1) is used to integrate genes of the vitamin B12 synthetic pathway into strain, which improves the vitamin B12 yield [45]. Conversely, *Shewanella oneidensis* is promising for bioenergy production, but its application is limited by substrate dependence on three-carbon compounds [54]. Cheng, *et al*. employ the improved *Sh*CAST to integrate two key genes to engineer a strain that can utilize six-carbon glucose and fructose [43].

Meanwhile, the type I-F CAST-based MUCICAT is applied to increase the copy numbers of the gene encoding glucose dehydrogenase (GDH) in *E. coli* BL21(DE3) to enhance GDH production [50]. MUCICAT is also exploited to optimize the N-acetylglucosamine (GlcNAc) biosynthesis in *E. coli* [55]. Multiplexed integration generates a library with 1–11 copies of the GlcNAc cassette. The strain harboring 5 copies of GlcNAc yields the highest GlcNAc titer [55].

Gene editing within a complex bacterial community

Understanding the functions of many bacteria and archaea is challenging due to the difficulty in culturing these microorganisms [56]. Achieving programmable organism- and locus-specific editing within a microbial community is desired but challenging, given that most microorganisms exist in communities [57]. Vo, *et al.* tackle this problem by conjugating the all-in-one

INTEGRATE plasmid from an *E. coli* into the mixed bacterial community, which achieves efficient RNA-guided transposition across distinct microbiome community sources [52]. Meanwhile, *Vc*DART (Table 1), coupled with ET-seq, facilitates selection-free tagging and tracking of genetic mutant fitness in synthetic soil communities. It allows simultaneous loss-of-function and gain-of-function mutations, exerting strong selective pressure for the enrichment of two bacteria in the community. Additionally, *Vc*DART achieves strain-specific editing in an infant gut consortium, enabling the enrichment and isolation of edited community members [53]. Consequently, CASTs offer a tool to investigate the function and fitness of specific genes in microbial communities, overcoming challenges associated with isolating unculturable bacteria. This capability suggests the potential use of CASTs in inhibiting methane-producing bacteria in the guts of ruminant livestock to mitigate methane production, a major greenhouse gas contributing to global warming.

Targeted mutagenesis and genetic screening in diverse bacteria

The *Sh*CAST-based STAGE (Table 1) is used to construct transposon mutant libraries, and further generate *P. aeruginosa* loss-of-function mutant libraries by interrupting 593 transcription factors (TFs) [58]. This method allows for identification of TFs for antibiotic resistance. Another study leverages *Sh*CAST for targeted mutagenesis in 3 classes of proteobacteria, especially in Betaproteobacteria and Gammaproteobacteria [59].

These studies altogether implicate the potentials of DSB-free CASTs to engineer prokaryotes that are difficult to edit, especially those in a mixed community, thus allowing for not only metabolic engineering and fundamental research, but also climate change control.

CASTs and OMEGA for Eukaryotic Genome Engineering

CASTs

Tou, *et al.* discover that a type V-K CAST from *Nostoc Sp.* PCC7101 (N_7CAST) enables DNA insertion in the HEK293 cell lysate [46]. The authors further construct N_7HELIX (Fig. 4A), which comprises pN_7HELIX to express transposition proteins, pTarget, pdonor, psgRNA, and pS15 expressing S15 (a host ribosomal protein required for type V-K CASTs, Box 1). Delivering the system into HEK293T cells enables insertion of a 2.6 kb cargo into the target plasmid 57-62 bp downstream of the PAM with only 7.9% co-integrate products, a 5.3-fold improvement compared with N_7CAST. However, the integration efficiency is very low (<0.1%) [46].

Recently Lampe, *et al.* construct pDonor, pTarget and plasmids expressing proteins and RNA components from a type I-E Cascade from *Pseudomonas sp*. S-6–2 (Fig. 4B) [60]. Delivering these plasmids into HEK293T cells enables cargo integration from pDonor into pTarget, at efficiencies <0.1%. To increase the efficiency, the authors screen 18 additional CASTs and identify a transposon from *Pseudoalteromonas* sp. (*Pse*CAST). Using the same approach and systemic engineering, the engineered *Pse*CAST confers 3-5% cargo transposition from pDonor to to pTarget, with the integration efficiency decreasing with cargo size (0.8 to15 kb). The authors further optimize *Pse*CAST and achieve donor integration into several genomic loci with efficiencies varying from 0.005-0.01%. The authors identify the essential roles of post-transposition complex (PTC) unfolding and disassembly for transposition. By co-expressing bacterial ClpX, an ATPase from MuA transposon that unfolds protein substrates, the integration efficiency is improved to ≈1%, indicating that PTC disassembly limits integration into genomic sites in HEK293T cells [60]. This work lays a foundation to exploit CASTs for eukaryotic genome engineering and demonstrates the possibility to improve the editing efficiency through systematic engineering and discovery of new

helper proteins. However, the editing efficiency is low compared with CRISPR-Cas9 and CRISPR-Cas12 systems. Furthermore, not all CASTs can edit eukaryotic genomes.

OMEGA

In 2021, Altae-Tran, *et al*. discover that prokaryotic IscB (≈400-500 aa) functions in conjunction with non-coding ωRNA to enable eukaryotic DNA cleavage [23]. ωRNA is encoded within the LE of the transposon and is longer than 200 nt. IscB-ωRNA cleaves dsDNA with a 3' target-adjacent motif (TAM) preference. Among 6 IscB proteins, a compact OgeuIscB (496 aa) induces indels in the genome of HEK293FT cells with varying efficiencies up to 4.4% [23]. Subsequent studies show that ωRNA plays the equivalent function of REC domains in Cas9 [38] and replaces some structural elements of Cas9 and crRNA/tracrRNA duplex [61].

In the meantime, Karvelis, *et al*. uncover that TnpB from *Deinococcus radiodurans* ISDra2 (408 aa) can complex with the ωRNA (≈150 nt) encoded within the RE of transposon [24]. TnpB-ωRNA complex recognizes the TAM and mediates *in vitro* plasmid cleavage. Delivering the plasmid encoding TnpB and ωRNA (Fig. 4C) into HEK293T cells induces indels at frequencies (10–20%) comparable to CRISPR–Cas9 and CRISPR-Cas12 [24]. ωRNA directs TnpB to create a staggered DNA cut in a manner similar to Cas12. TnpB represents the minimal structural and functional core of Cas12 and is the evolutionary ancestor of Cas12 [62]. TnpB can processes its own mRNA into ωRNA [63], which is conserved among all guide RNAs of Cas12 [64].

Recently, Xiang *et al.* identify new TnpB proteins, particularly ISAam1 (369 aa) and ISYmu1 (382 aa), and showcase their ability to induce indels across genomic loci in HEK293T cells [65]. ISAam1 and ISYmu1 outperform 3 Cas12f editors, exhibiting editing efficiency comparable to SaCas9 (1,053 aa) but with a significantly smaller size. They predominantly induce deletions, and insertions are less common, in a way similar to those in CRISPR-Cas systems. Notably, the

compact ISAam1 editor is packed into AAV8 and injected into mice, achieving >5% editing efficiency that outperforms Un1Cas12f1 and SaCas9. This groundbreaking study demonstrates the *in vivo* delivery of the hypercompact OMEGA system via AAV and establishes TnpB's potential as a human genome editing tool.

Fanzor (Fz) protein is a class of endonucleases encoded by transposons discovered in various eukaryotes ranging from fungi to flies and even eukaryotic viruses [35, 66]. Fz is categorized into Fz1 and Fz2, both arising from prokaryotic IS607 TnpBs [67] and can coordinate with ωRNA for RNA-guided genome cleavage in human cells (Fig. 4D). Fz is also smaller than Cas12a nucleases, an important feature for genome editing applications [35]. Saito*, et al.* screen multiple Fz homologs and show that 3 Fz homologs can induce indels at genomic sites in HEK293T cells with efficiencies ranging from 0.01% to 11.8%. The overall editing efficiency of 3 Fz variants is comparable with that of a minimal Cas12 editor, AsCas12f1 (422 aa) [14]. Two of the homologs cause large deletion, with a pattern similar to that of Cas12a. Mutating SpuFz1 (638 aa) and optimizing the ωRNA architecture increase indel mutation frequency (>10%) at multiple loci. SpuFz1 efficiently performs programmable RNA-guided DNA cleavage, matching the effectiveness of miniature AsCas12f1. In contrast to TnpB and Cas12, all examined Fn orthologs exhibit no collateral cleavage activity on dsDNA, dsRNA, ssDNA, or ssRNA when bound to target dsDNA [39].

Jiang, *et al*. further reveal that Fz proteins encode a nuclear localization signal (NLS) essential for nuclear transport to access genome [68]. The intrinsic NLS obviates the need to fuse extra NLS, rendering Fz advantageous to other prokaryotic OMEGA nucleases for eukaryotic genome editing. These Fz orthologs enable primarily large deletions with varying editing frequencies (0.5-15%) at different loci.

These studies validate the functionality of Fz proteins as eukaryotic OMEGA RNA-guided nucleases in mammalian cells. The creation of staggered ends by Fz proteins would allow for specific DNA integration through homology-directed repair and enable extensive sequence deletions for knockout mutagenesis. This approach enhances the comprehension of genetic and epigenetic landscapes in the genome. However, Fz proteins exhibit lower DNA cleavage efficiencies compared to CRISPR-Cas9 and -Cas12, and they have not been employed for the integration of foreign genes.

Concluding Remarks and Future Perspectives

CRISPR-Cas system is not suitable for gene integration into prokaryotes with poor DNA repair machinery or in a complex community. The DSB-free property of CASTs makes them ideal to integrate large genetic payload into prokaryotes that are difficult to engineer using conventional CRISPR systems [20, 21, 25-28, 30] and bacteria in a mixed community for various applications. Note, however, that CASTs-mediated gene integration is not scarless. Caution should be used when editing is performed in the operons containing gene clusters.

Both type V-K and I-F CASTs are burgeoning tools for eukaryotic genome editing, but their editing efficiencies are still too low. Compared with other emerging DSB-free prime editors (e.g. TwinPE [69], PASTE [70]) and dCas9 fusion proteins [71, 72] which can integrate large genetic payload with higher efficiencies, CASTs are currently less practical. Despite the low efficiency, this field has captured growing interests. The inefficient editing of eukaryotic genomes by the prokaryotes-derived CASTs may be attributed to the alteration of transposon protein structures in eukaryotic cells, difficulty to optimize the relative expression levels of individual CASTs proteins and the requirements of CASTs for host-encoded factors to facilitate transposition. Future bioinformatic mining to search appropriate CASTs (or even eukaryotic CASTs) for eukaryotic

engineering is necessary. Further research is needed to identify eukaryotic host proteins that can collaborate with CASTs components (see Outstanding Questions), clarify the protein structure and transposition mechanisms, and engineer individual proteins to enhance transpososome assembly in eukaryotes. Once these mechanisms and factors are understood and manipulated, CASTs have the potential to become a novel category of DSB-free genome editors for eukaryotes.

To date, the large size of most Cas9 and Cas12 remain the bottleneck for their *in vivo* delivery by AAV [11, 12]. Although miniature Cas12f homologs are active in human cells, the overall editing efficiency is below 10% [13-15]. Therefore, the small OMEGA IscB and TnpB endonucleases discovered in prokaryotic transposons provide promising new toolkits for eukaryotic genome editing. Specifically, hypercompact TnpB orthologs demonstrate superior performance to Cas12f editors *in vitro* and outperform both Un1Cas12f1 and SaCas9 *in vivo* [65]. TnpB constitutes the largest group of genes in prokaryotes, with more than one million putative loci [73]. The enormous diversity of TnpBs makes it promising to discover more miniature genome editors for facile delivery and eukaryotic genome engineering. Yet whether bacterial TnpB elicits immune responses and genome aberrations in mammals like Cas9/Cas12 remain to be elucidated.

The compact Fz protein derived from eukaryotic transposons provides another exciting option for evading immune responses, especially if human Fz homologs can be identified (see Outstanding Questions). The broad diversity of Fz across eukaryotes suggests the potential discovery of additional Fz endonucleases, expanding the genome engineering toolbox. While current achievements in mammalian cells show low indel mutation rates and no transgene integration, there is room for optimization of Fz components to enhance editing efficiency [39]. The eukaryotic origin, structural attributes, and miniature size of Fz proteins position them as potential alternatives to large CRISPR/Cas endonucleases for human genome editing and unexplored applications in

synthetic and applied biology. Note that both TnpB and Fz endonuclease families necessitate more complex TAM sequences compared to the PAM sequences of the CRISPR-Cas12 system. Engineering the domain for TAM recognition is essential to expand the targeting range [32]. Furthermore, whether TnpB and Fz proteins elicit immune responses require further investigations (see Outstanding Questions).

In conclusion, CASTs and OMEGA are burgeoning genome engineering tools derived from transposons. Albeit in their infancy, there is growing enthusiasm to explore CASTs and OMEGA systems, indicating a paradigm shift towards the use of transposons for genome editing.

Glossary

AAV: adeno-associated virus, the most-widely used gene delivery vector and is approved by Food and Drug Administration (FDA). However, its packaging capacity is only ≈4.7 kb [11, 12], rendering co-delivery of Cas proteins, genetic cargo and other transcriptional regulators challenging.

Cas9: An endonuclease containing an HNH domain that nicks the target DNA strand, and a RuvC domain that nicks the non-target strand.

crRNA: CRISPR RNA that encodes a spacer sequence complementary to the target DNA.

DNA repair: In eukaryotic cells, DSB is repaired through non-homologous end joining (NHEJ) pathway if a homologous DNA is absent, resulting in insertions/deletions (indels). With a template DNA carrying the sequence homologous to the targeted locus, DSB can be repaired by homology-directed repair (HDR) pathway, leading to foreign gene integration [27, 74].

gRNA: A simplified artificial chimeric guide RNA by fusing crRNA and tracrRNA [4-6].

insertion sequence (IS): Widespread transposons among which the IS200/IS605 families are the simplest [75]. IS200/IS605 families perform transposition through a single-strand 'peel and paste' mechanism [76]. These IS elements encode a well-documented TnpA transposase [77] and often encode an accessory TnpB nuclease. The well-characterized *Deinococcus radiodurans* ISDra2 of the IS200/IS605 family comprises overlapping *tnpA* and *tnpB* genes flanked by LE and RE elements [24].

PAM: protospacer-adjacent motif, the DNA sequence that can be recognized by the Cas9/crRNA/tracrRNA complex for the cleavage of adjacent sequence. Different Cas proteins hava different PAM preferences.

tracrRNA: small transacting RNA partially base pairing with a pre-crRNA to form a crRNA/tracrRNA hybrid duplex. Cas9 complexes with the crRNA/tracrRNA to coordinate the recognition of protospacer-adjacent motif on the target DNA and trigger DSB [11].

transposons: mobile genetic elements (MGEs) that can move from one location to another within a host genome and are widespread in prokaryotes, eukaryotes and even viruses [28]. The prototypic Tn7 transposon from *E. coli* contains characteristic left end (LE) and right end (RE) sequences, cargo gene and encodes five intervening genes (tnsA to tnsE, Fig. 1A). The core transposase proteins, TnsA, TnsB, and TnsC direct transposition in two modes: homing and mobile element transposition. In the homing mode, the heteromeric TnsABC complex interacts with the DNA-binding protein TnsD for site-specific transposition into the chromosomal homing site [78]. In the second mode, TnsABC interact with TnsE to direct the transposon preferentially to mobile elements such as conjugative plasmids [47, 79].

Box

Box 1. Mechanisms of type V-K CASTs transposition: In type V-K CASTs, the overall structure of Cas12k resembles other Cas12 proteins. Cryo-electron microscopy (cryo-EM) studies show that Cas12k adopts a bi-lobed structure connected by a loop [80, 81]. The N-terminal lobe is composed of the wedge (WED), recognition (REC) and PAM interacting (PI) domains, in which WED domain is crucial for recognizing crRNA/tracrRNA. The tracrRNA serves as a scaffold to position Cas12k and crRNA for pinpointing complementary PAM sequences [80, 81]. Upon binding to a 5'-GTN-3' PAM, Cas12k initiates RNA hybridization to the target strand DNA, hence inducing incomplete R-loop formation [82]. TniQ recognizes tracrRNA and R-loop to bridge tracrRNA and TnsC filament [82].

Conversely, TnsC prefers to form helical filaments unidirectionally on AT-rich DNA stretches in the presence of ATP [83]. In the next step, TniQ contacts TnsC to prime the polymerization of TnsC filament along the target DNA and terminate the TnsC polymerization [44, 82-85]. The TnsC filaments interact with one DNA strand within the duplex, providing a target site search mechanism to define downstream insertion polarity [83]. The complex establishes the connection between the CRISPR module and the transposition module (Fig. 2B).

In the last step, TnsC filaments recruit the tetrameric TnsB transposase [85], which triggers TnsC filaments disassembly upon ATP hydrolysis to expose the insertion site and then TnsB catalyzes donor DNA transposition [81, 82, 84]. TnsB interacts with TnsC filaments on only one face, which stimulates the ATPase activity [84, 85]. The C-terminal end of TnsB adopts a short, structured 15-residue hook that decorates TnsC filaments, hence serving a pivotal role in transposase recruitment to the target site [85]. TnsB functions as a site-specific 3'-5' exonuclease, which specifically cleaves the transferred strand of donor DNA at the junction of the flanking DNA and terminal repeats [86]. The interactions between TnsB and TnsC directs DNA insertion in a fixed position relative to the target DNA recognition site of the CRISPR module (Fig. 2B). These studies demonstrate that the transpososome, consisting of crRNA/tracrRNA-bound Cas12k, one TniQ subunit, one ribosomal S15 subunit, four subunits of TnsB, and two full turns of TnsC with six subunits per turn [84], cooperatively directs the programmable DNA transposition (Fig. 2B).

One problem for type V-K CASTs is off-target integration [17]. Saito and coworkers discover that type V-K CASTs preferentially localize to tRNA genes and mediate transposition with two different mechanisms: (i) crRNA-guided transposition and (ii) crRNA-independent homing [49]. The homing to off-target site is guided by a short, delocalized crRNA. In agreement, George, et al. unveil that type V-K CASTs undergo RNA-dependent and RNA-independent transposition [44].

RNA-dependent transposition relies on Cas12k for accurate target selection and TnsB preferentially integrates into sites within the target site duplication [44, 87].

These studies indicate that the stoichiometry of type V-K CAST proteins governs the integration efficiency and specificity. Raising the expressing levels of Cas12k alone [42] or the entire TnsC/TnsB/TniQ/Cas12k cassette [81, 83] increases the integration efficiency. High TnsC expression decreases the accuracy, while low TnsC expression enhances the on-target integration frequency to 98% [44], indicating that TnsC filament assembly drives RNA-independent, untargeted transposition [44]. Conversely, TnsB disassembles TnsC filaments [83, 85], rescuing the specificity at on-target sites when TnsC concentration is high [44]. These data implicate that increasing TnsB expression and lowering TnsC determine the transposition pathway choice and enhance on-target integration [44].

Box 2. Mechanisms of type I CASTs transposition: In type I-F3 *Vch*CAST, one Cas6 subunit, six subunits of Cas7 and one Cas8/5 molecule form the Cascade complex and associate with crRNA [18, 88]. Cascade binds DNA and TniQ, recruits TnsC and the heteromeric TnsAB transposase in complex with the donor DNA, to assemble the catalytically active transpososome (Fig. 3B) [87]. The Cas8/5 fusion protein binds the 5' crRNA handle and contacts the TniQ dimer [18]. The dimeric TniQ binds to the Cascade complex in a head-to-tail configuration at the interface formed by Cas6 and Cas7 near the 3' end of crRNA [89]. One TniQ interacts with Cas6 whereas the other TniQ interacts with Cas7 [88]. The Cascade-TniQ complex results in the complete R-loop formation and scans numerous sites in the target DNA to specify the site of DNA integration [20]. Notably, TniQ-Cascade complex alone binds not only at the target site specified by Cascade's crRNA, but also many off-target sites with mismatches [90]. Most off-target sites fail to recruit

TnsC and TnsB to form catalytically active transpososome. Only a subset is bound by TnsC and an even smaller subset is bound by TnsB, hence minimizing off-target integration [90].

Type I-F TnsC bridges the RNA-guided DNA targeting module (TniQ-Cascade) and the DNA integration module (TnsAB), hence contributing to the integration fidelity [91]. In the presence of ATP, TnsC forms a heptameric ring architecture with a central pore that threads DNA like a needle, but does not appear to form higher-order filaments [90] as observed in type V-K CASTs. The TnsC heptameric ring recruits the TnsAB-loaded donor DNA and subsequently integrates the donor DNA cargo at the location usually 47–51 bp downstream of the Cascade target site flanked by a 5-bp target site duplication (Fig. 3B).

Similar to CRISPR-Cas systems, the crRNA/tracrRNA or crRNA in CASTs can be engineered as the chimeric gRNA carrying the guide sequence for PAM recognition [92]. CASTs systems are more complex than the class 2 CRISPR-Cas systems, but do not require DSB induction. Both type V-K and I-F CASTs encode a core transposition machinery composed of a transposase TnsB for integration, a AAA+ ATPase TnsC for target site selection, and an adaptor protein TniQ [17, 18, 22, 78]. Despite the similarities, differences between these two CASTs exist [21, 26, 30]. Type V-K TnsC forms continuous helical filaments on DNA [84], whereas type I-F3 TnsC forms heptameric structure [90]. Moreover, type V-K CASTs require host ribosomal protein S15 while type I CASTs require a host-encoded integration host factor (IHF).

Figure Captions

Figure 1. CASTs and OMEGA. (A) CASTs. Both type V-K and type I CASTs are Tn7-like transposons but are linked to CRISPR. Type V-K CASTs encode four proteins (TnsB, TnsC, TniQ, Cas12k), tracrRNA/crRNA and harbor the genetic cargo within the LE and RE ends. Type I-F3

CASTs encode seven proteins (TnsA, TnsB, TnsC, Cas6, Tns7, a fusion protein Cas5/8 comprising Cas8 and Cas5, and TniQ) and expresses only crRNA (without tracrRNA). (B) OMEGA. IscB and TnpB are encoded in the IS200/IS605 family transposons. IscB and TnpB are evolutionary ancestors of Cas9 and Cas12, respectively. IscB shares HNH and RuvC nuclease domains with Cas9 proteins and associates with OMEGA RNA (ωRNA) to recognize the RNA-targeted DNA by transposon-associated motive (TAM), enabling IscB to cleave its target DNA using the HNH and RuvC domains. Similarly, TnpB associates with the ωRNA guide for dsDNA cleavage.

Figure 2. Development and transposition mechanisms of type V-K CASTs. (A) Original ShCAST, SHOT and HELIX. HELIX is engineered from ShCAST by fusing TnsB to a mutated homing endonuclease nAnil. (B) The detailed transposition procedures. The transpososome consists of gRNA-bound Cas12k, one TniQ subunit, four subunits of TnsB, two full turns of TnsC with six subunits per turn and a host-encoded ribosomal protein S15. Cas12k acts in concert with gRNA to bind PAM sequence without triggering DSB, and recruits the transposase complex to integrate DNA cargo into the target site ≈60-66 bp downstream of the PAM. TSD, target site duplication.

Figure 3. Development and transposition mechanisms of type I CASTs. (A) Original INTEGRATE system containing 3 plasmids and the streamlined all-in-one INTEGRATE. (B) The detailed transposition procedures. The INTEGRATE system recognizes a flexible PAM, and achieves transposition in either forward or reverse direction with a 5-bp TSD. TniQ forms essential interactions with Cascade (composed of Cas6, Cas7 and Cas8/5). Cascade binds DNA and TniQ, recruits TnsC and the heteromeric TnsAB transposase in complex with the donor DNA, to assemble the catalytically active transpososome. The transpososome subsequently integrates the donor DNA cargo at the location downstream of the Cascade target site flanked by a 5-bp TSD.

Figure 4. CASTs and OMEGA for genome engineering in eukaryotes. (A) Type V-K CAST. The N_7HELIX enables insertion of a 2.6 kb into the target plasmid 57-62 bp downstream of the PAM in the HEK293T cells. The proteins are fused with bipartite (BP) NLS and may be linked by 2A peptides. (B) Type I CAST. Co-transfection of pDonor harboring the mini-transposon, pTarget and plasmids expressing proteins and RNA components into HEK293T cells enables cargo integration from the pDonor plasmid into pTarget. (C) OMEGA-encoded TnpB for genome engineering. The plasmid encoding TnpB and ωRNA optimized for expression in eukaryotic cells is transfected into HEK293T cells and induces mutation. (D) Fn-mediated genome editing in HEK293FT cells. Transfection of plasmids encoding Fz and ωRNA induces RNA-guided DNA cleavage in the genome of human cells.

Competing interests

The authors declare no competing interests.

Author contributions

Chang, C.-W., Pham, N.M. and Truong, A.V draw the figures, prepare tables and write the paper. Hu, Y.-C. supervises the project and writes the paper.

Acknowledgement

The authors acknowledge the financial support from the National Science and Technology Council (NSTC 112-2223-E-007-002, 112-2314-B-007-004-MY3, 111-2223-E-007-003, 111-2634-F-007-007, 111-2622-E-007-003, 111-2622-8-007-011) and National Health Research Institutes (NHRI-EX112-11014BI, NHRI-EX113-11329EI), Taiwan.

References

1. Makarova, K.S. *et al.* (2020) Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. Nat. Rev. Microbiol. 18, 67-83.
2. Cong, L. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819-23.
3. Mali, P. *et al.* (2013) RNA-guided human genome engineering via Cas9. Science 339, 823-826.
4. Jinek, M. *et al.* (2012) A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. Science 337, 816-821.
5. Jiang, W., Bikard, D., Cox, D., Zhang, F. and Marraffini, L.A. (2013) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat. Biotechnol. 31, 233-239.
6. Arnoult, N. *et al.* (2017) Regulation of DNA repair pathway choice in S and G2 phases by the NHEJ inhibitor CYREN. Nature 549, 548-552.
7. Chung, M.E. *et al.* (2017) Enhanced integration of large DNA into *E. coli* chromosome by CRISPR/Cas9. Biotechnol. Bioeng. 114, 172-183.
8. Kim, S. *et al.* (2018) CRISPR RNAs trigger innate immune responses in human cells. Genome Res. 28, 367-373.
9. Adikusuma, F. *et al.* (2018) Large deletions induced by Cas9 cleavage. Nature 560, E8-E9.
10. Kosicki, M., Tomberg, K. and Bradley, A. (2018) Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. Nat Biotechnol 36, 765-771.
11. Doudna, J.A. (2020) The promise and challenge of therapeutic genome editing. Nature 578, 229-236.
12. Rousset, F. and Sorek, R. (2021) A treasure trove of molecular scissors. Science 374, 37-38.

13. Kim, D.Y. *et al.* (2022) Efficient CRISPR editing with a hypercompact Cas12f1 and engineered guide RNAs delivered by adeno-associated virus. Nat. Biotechnol. 40, 94-102.

14. Wu, Z. *et al.* (2021) Programmed genome editing by a miniature CRISPR-Cas12f nuclease. Nat Chem Biol 17, 1132-1138.

15. Xu, X. *et al.* (2021) Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing. Mol. Cell 81, 4333-4345.e4.

16. Peters, J.E., Makarova, K.S., Shmakov, S. and Koonin, E.V. (2017) Recruitment of CRISPR-Cas systems by Tn7-like transposons. Proc. Nat. Acad. Sci. 114, E7358-E7366.

17. Strecker, J. *et al.* (2019) RNA-guided DNA insertion with CRISPR-associated transposases. Science 365, 48-53.

18. Klompe, S.E., Vo, P.L.H., Halpin-Healy, T.S. and Sternberg, S.H. (2019) Transposon-encoded CRISPR–Cas systems direct RNA-guided DNA integration. Nature 571, 219-225.

19. Klompe, S.E. *et al.* (2022) Evolutionary and mechanistic diversity of Type I-F CRISPR-associated transposons. Mol. Cell 82, 616-628.e5.

20. Koonin, E.V. and Krupovic, M. (2023) New faces of prokaryotic mobile genetic elements: Guide RNAs link transposition with host defense mechanisms. Curr. Opin. Syst. Biol. 36, 100473.

21. Tenjo-Castaño, F., Montoya, G. and Carabias, A. (2022) Transposons and CRISPR: Rewiring gene editing. Biochemistry.

22. Hsieh, S.C. and Peters, J.E. (2023) Discovery and characterization of novel type I-D CRISPR-guided transposons identified among diverse Tn7-like elements in cyanobacteria. Nucleic Acids Res 51, 765-782.

548 23. Altae-Tran, H. *et al.* (2021) The widespread IS200/IS605 transposon family encodes diverse
549 programmable RNA-guided endonucleases. Science 374, 57-65.
550 24. Karvelis, T. *et al.* (2021) Transposon-associated TnpB is a programmable RNA-guided DNA
551 endonuclease. Nature 599, 692-696.
552 25. Volke, D.C., Orsi, E. and Nikel, P.I. (2023) Emergent CRISPR–Cas-based technologies for
553 engineering non-model bacteria. Curr. Opin Microbiol. 75, 102353.
554 26. Liu, Z. *et al.* (2023) Endogenous CRISPR-Cas mediated in situ genome editing: State-of-the-
555 art and the road ahead for engineering prokaryotes. Biotechnol. Adv. 68, 108241.
556 27. Nambiar, T.S., Baudrier, L., Billon, P. and Ciccia, A. (2022) CRISPR-based genome editing
557 through the lens of DNA repair. Mol. Cell 82, 348-388.
558 28. Wang, X. *et al.* (2023) Long sequence insertion via CRISPR/Cas gene-editing with transposase,
559 recombinase, and integrase. Curr. Opin. Biomed. Eng. 28, 100491.
560 29. Alalmaie, A., Diaf, S. and Khashan, R. (2023) Insight into the molecular mechanism of the
561 transposon-encoded type I-F CRISPR-Cas system. Journal of Genetic Engineering and
562 Biotechnology 21, 60.
563 30. Tou, C.J. and Kleinstiver, B.P. (2022) Recent advances in double-strand break-free kilobase-
564 scale genome editing technologies. Biochemistry.
565 31. Ma, W., Xu, Y.-S., Sun, X.-M. and Huang, H. (2021) Transposon-associated CRISPR-Cas
566 system: A powerful DNA insertion tool. Trends Microbiol 29, 565-568.
567 32. Badon, I.W., Oh, Y., Kim, H.-J. and Lee, S.H. (2024) Recent application of CRISPR-Cas12
568 and OMEGA system for genome editing. Molecular Therapy 32, 32-43.
569 33. Awan, M.J.A., Awan, M.R.A., Amin, I. and Mansoor, S. (2023) Fanzor: a compact
570 programmable RNA-guided endonuclease from eukaryotes. Trends Biotechnol 41, 1332-1334.

571 34. Yang, H. and Patel, D.J. (2023) Fanzors: Striking expansion of RNA-guided endonucleases to
572 eukaryotes. Cell Research.
573 35. Patinios, C. and Beisel, C.L. (2023) For the CRISPR Fan(zor)atics: RNA-guided DNA
574 endonucleases discovered in eukaryotes. Mol Cell 83, 3046-3048.
575 36. Rybarski, J.R., Hu, K., Hill, A.M., Wilke, C.O. and Finkelstein, I.J. (2021) Metagenomic
576 discovery of CRISPR-associated transposons. Proc Natl Acad Sci U S A 118.
577 37. Meers, C. *et al.* (2023) Transposon-encoded nucleases use guide RNAs to promote their selfish
578 spread. Nature 622, 863-871.
579 38. Schuler, G., Hu, C. and Ke, A. (2022) Structural basis for RNA-guided DNA cleavage by IscB-
580 ωRNA and mechanistic comparison with Cas9. Science 376, 1476-1481.
581 39. Saito, M. *et al.* (2023) Fanzor is a eukaryotic programmable RNA-guided endonuclease. Nature
582 620, 660-668.
583 40. Strecker, J., Ladha, A., Makarova, K.S., Koonin, E.V. and Zhang, F. (2020) Response to
584 Comment on “RNA-guided DNA insertion with CRISPR-associated transposases”. Science
585 368, eabb2920.
586 41. Rice, P.A., Craig, N.L. and Dyda, F. (2020) Comment on "RNA-guided DNA insertion with
587 CRISPR-associated transposases". Science 368.
588 42. Chang, C.-W. *et al.* (2023) Metabolic engineering of *Escherichia coli* to enhance protein
589 production by coupling ShCAST-based optimized transposon system and CRISPR
590 interference. J. Taiwan Ins. Chem. Eng. 144, 104746.
591 43. Cheng, Z.H. *et al.* (2022) Repurposing CRISPR RNA-guided integrases system for one-step,
592 efficient genomic integration of ultra-long DNA sequences. Nucleic Acids Res. 50, 7739-7750.

593 44. George, J.T. *et al.* (2023) Mechanism of target site selection by type V-K CRISPR-associated
594 transposases. Science 382, eadj8543.
595 45. Cui, Y. *et al.* (2022) A versatile Cas12k-based genetic engineering toolkit (C12KGET) for
596 metabolic engineering in genetic manipulation-deprived strains. Nucleic Acids Res. 50, 8961-
597 8973.
598 46. Tou, C.J., Orr, B. and Kleinstiver, B.P. (2023) Precise cut-and-paste DNA insertion using
599 engineered type V-K CRISPR-associated transposases. Nat. Biotechnol.
600 47. Peters, J.E. (2019) Targeted transposition with Tn7 elements: safe sites, mobile plasmids,
601 CRISPR/Cas and beyond. Mol Microbiol 112, 1635-1644.
602 48. Wimmer, F., Mougiakos, I., Englert, F. and Beisel, C.L. (2022) Rapid cell-free characterization
603 of multi-subunit CRISPR effectors and transposons. Mol. Cell 82, 1210-1224.e6.
604 49. Saito, M. *et al.* (2021) Dual modes of CRISPR-associated transposon homing. Cell 184, 2441-
605 2453.e18.
606 50. Zhang, Y. *et al.* (2020) Multicopy chromosomal integration using CRISPR-associated
607 transposases. ACS Synth. Biol. 9, 1998-2008.
608 51. Yang, S.Q. *et al.* (2021) Orthogonal CRISPR-associated transposases for parallel and
609 multiplexed chromosomal integration. Nucleic Acids Res. 49, 10192-10202.
610 52. Vo, P.L.H. *et al.* (2021) CRISPR RNA-guided integrases for high-efficiency, multiplexed
611 bacterial genome engineering. Nat Biotechnol 39, 480-489.
612 53. Rubin, B.E. *et al.* (2022) Species- and site-specific genome editing in complex bacterial
613 communities. Nat. Microbiol 7, 34-47.

54. Cheng, L. *et al.* (2020) Developing a base-editing system to expand the carbon source utilization spectra of Shewanella oneidensis MR-1 for enhanced pollutant degradation. Biotechnol Bioeng 117, 2389-2400.

55. Zhang, Y. *et al.* (2021) Programming cells by multicopy chromosomal integration using CRISPR-associated transposases. CRISPR J 4, 350-359.

56. Jansson, J.K., McClure, R. and Egbert, R.G. (2023) Soil microbiome engineering for sustainability in a changing environment. Nat. Biotechnol. 41, 1716-1728.

57. Sheth, R.U., Cabral, V., Chen, S.P. and Wang, H.H. (2016) Manipulating Bacterial Communities by in situ Microbiome Engineering. Trends Genet 32, 189-200.

58. Chen, W.Z. *et al.* (2021) Targeted genetic screening in bacteria with a Cas12k-guided transposase. Cell Reports 36.

59. Rodriguez, L.T., Ellington, A.J., Reisch, C.R. and Chevrette, M.G. (2023) CRISPR-associated transposase for targeted mutagenesis in diverse proteobacteria. ACS Synth. Biol. 12, 1989-2003.

60. Lampe, G.D. *et al.* (2024) Targeted DNA integration in human cells without double-strand breaks using CRISPR-associated transposases. Nat Biotechnol 42, 87-98.

61. Kato, K. *et al.* (2022) Structure of the IscB-ωRNA ribonucleoprotein complex, the likely ancestor of CRISPR-Cas9. Nat Commun 13, 6719.

62. Sasnauskas, G. *et al.* (2023) TnpB structure reveals minimal functional core of Cas12 nuclease family. Nature 616, 384-389.

63. Nety, S.P. *et al.* (2023) The transposon-encoded protein TnpB processes its own mRNA into ωRNA for guided nuclease activity. Crispr J 6, 232-242.

64. Nakagawa, R. *et al.* (2023) Cryo-EM structure of the transposon-associated TnpB enzyme. Nature 616, 390-397.

65. Xiang, G. *et al.* (2023) Evolutionary mining and functional characterization of TnpB nucleases identify efficient miniature genome editors. Nat. Biotechnol.

66. Bao, W. and Jurka, J. (2013) Homologues of bacterial TnpB_IS605 are widespread in diverse eukaryotic transposable elements. Mobile DNA 4, 12.

67. Yoon, P.H. *et al.* (2023) Eukaryotic RNA-guided endonucleases evolved from a unique clade of bacterial enzymes. Nucleic Acids Res. 51, 12414-12427.

68. Jiang, K. *et al.* (2023) Programmable RNA-guided DNA endonucleases are widespread in eukaryotes and their viruses. Sci. Adv. 9, eadk0171.

69. Anzalone, A.V. *et al.* (2022) Programmable deletion, replacement, integration and inversion of large DNA sequences with twin prime editing. Nat. Biotechnol. 40, 731-740.

70. Yarnall, M.T.N. *et al.* (2023) Drag-and-drop genome insertion of large sequences without double-strand DNA cleavage using CRISPR-directed integrases. Nat. Biotechnol. 41, 500-512.

71. Pallarès-Masmitjà, M. *et al.* (2021) Find and cut-and-transfer (FiCAT) mammalian genome engineering. Nat. Commun 12.

72. Wang, C. *et al.* (2022) dCas9-based gene editing for cleavage-free genomic knock-in of long sequences. Nat. Cell Biol. 24, 268-278.

73. Altae-Tran, H. *et al.* (2023) Diversity, evolution, and classification of the RNA-guided nucleases TnpB and Cas12. Proc. Natl. Acad. Sci. 120, e2308224120.

74. Hsu, M.-N. *et al.* (2019) CRISPR technologies for stem cell engineering and regenerative medicine. Biotechnol. Adv. 37, 107447.

658 75. Siguier, P., Gourbeyre, E. and Chandler, M. (2014) Bacterial insertion sequences: their genomic
659 impact and diversity. FEMS Microbiol. Rev. 38, 865-891.
660 76. He, S. *et al.* (2015) The IS200/IS605 family and "Peel and Paste" single-strand transposition
661 mechanism. Microbiol Spectr 3, 10.1128.
662 77. Barabas, O. *et al.* (2008) Mechanism of IS200/IS605 family DNA transposases: Activation and
663 transposon-directed target site selection. Cell 132, 208-220.
664 78. Faure, G. *et al.* (2023) Modularity and diversity of target selectors in Tn7 transposons.
665 Molecular Cell 83, 2122-2136.e10.
666 79. Peters, P.E. (2014) Tn7. Microbiol. Spect. 2, 2.5.27.
667 80. Xiao, R. *et al.* (2021) Structural basis of target DNA recognition by CRISPR-Cas12k for RNA-
668 guided DNA transposition. Mol Cell 81, 4457-4466.e5.
669 81. Querques, I., Schmitz, M., Oberli, S., Chanez, C. and Jinek, M. (2021) Target site selection and
670 remodelling by type V CRISPR-transposon systems. Nature 599, 497-502.
671 82. Schmitz, M., Querques, I., Oberli, S., Chanez, C. and Jinek, M. (2022) Structural basis for the
672 assembly of the type V CRISPR-associated transposon complex. Cell 185, 4999-5010.e17.
673 83. Park, J.U. *et al.* (2021) Structural basis for target site selection in RNA-guided DNA
674 transposition systems. Science 373, 768-774.
675 84. Park, J.-U. *et al.* (2023) Structures of the holo CRISPR RNA-guided transposon integration
676 complex. Nature 613, 775-782.
677 85. Park, J.-U., Tsai, A.W.-L., Chen, T.H., Peters, J.E. and Kellogg, E.H. (2022) Mechanistic
678 details of CRISPR-associated transposon recruitment and integration revealed by cryo-EM.
679 Proc. Natl. Acad. Sci. 119, e2202590119.

680 86. Zeng, T. *et al.* (2023) Mechanistic insights into transposon cleavage and integration by TnsB
681 of ShCAST system. Cell Rep. 42.
682 87. Walker, Matt W.G., Klompe, Sanne E., Zhang, Dennis J. and Sternberg, Samuel H. (2023)
683 Novel molecular requirements for CRISPR RNA-guided transposition. Nucleic Acids Res. 51,
684 4519-4535.
685 88. Jia, N., Xie, W., de la Cruz, M.J., Eng, E.T. and Patel, D.J. (2020) Structure–function insights
686 into the initial step of DNA integration by a CRISPR–Cas–Transposon complex. Cell Res. 30,
687 182-184.
688 89. Halpin-Healy, T.S., Klompe, S.E., Sternberg, S.H. and Fernandez, I.S. (2020) Structural basis
689 of DNA targeting by a transposon-encoded CRISPR-Cas system. Nature 577, 271-+.
690 90. Hoffmann, F.T. *et al.* (2022) Selective TnsC recruitment enhances the fidelity of RNA-guided
691 transposition. Nature 609, 384-393.
692 91. Hossain, A.A. and Marraffini, L.A. (2022) CRISPR-assisted transposition: TnsC finds (and
693 threads) the needle in the haystack. Mol Cell 82, 3968-3969.
694 92. Hou, Z. and Zhang, Y. (2019) Inserting DNA with CRISPR. Science 365, 25-26.
695

696 **Table 1. Development of CASTs for prokaryotic genome engineering**

697

	System	Expression features	Engineered strain	editing features	Ref
Type V-K	Original *Sh*CAST	pHelper carries a lac promoter-driven operon comprising tnsB, tnsC, tniQ, cas12k, together with gRNA. pDonor contains the genetic cargo flanked by the transposon LE/RE.	*E. coli* Pir	● integrates cargo into ≈60% of selected genomic sites with efficiencies varying from <10% to ≈80% ● integrates a cargo up to 10 kb.	[17]
	Improved *Sh*CAST	The *lac* promoter in the pHelper plasmid is swapped with promoters of different strengths to drive the tnsB-tnsC-tniQ-Cas12k-gRNA expression.	*Shewanella oneidensis* MR-1	● inserts a 2 kb cargo into 4 identical repeats at ≈100% efficiency. ● enables insertion of DNA up to 31.7 kb into genome at ≈100% efficiency in a single orientation	[43]
	SHOT (*Sh*CAST-based Optimized Transposon)	● increasing Cas12k expression, independent of TnsC/TnsB/TniQ, under a stronger promoter such as T7 or Tac. ● Decreasing the editing temperature from 37°C to 30°C and re-streaking the colonies.	*E. coli* BL21(DE3), MG1655, W3110, W, BW25113 strains and *Pseudomonas putida*	● enhances on-target integration efficiency to >90% in different bacteria strains. ● alleviates off-target effects for cargo gene into 3 tested loci and enables on-target integration of 14.5 DNA cargo with ≈100% efficiency	[42]
	C12KGET (Cas12k-based genetic engineering toolkit)	stronger J23119 promoter to improve the expression of transposase and Cas12k.	*Sinorhizobium meliloti*	● achieves genomic integration of fragments up to 10 kb in size with up to 100% efficiency	[45]
	HELIX (Homing Endonuclease -assisted Large-sequence Integrating	fusing nAnil to the N terminus of TnsB.	*E. coli* pir1, *Scytonema hofmannii*	● enables cut-and-paste DNA insertion with up to 99.4% simple insertion product purity, while retaining robust integration efficiencies on genomic targets. ● enables integration of DNA at least 9.8 kb into target genomic site with comparably high efficiency.	[46]

	CAST compleX)			• C-terminal fusions of one or two copies of TniQ to Cas12k retains up to 62-69% integration efficiency relative to that of unfused *Sh*CAST, while fusion of either TniQ or TnsC to the N-terminus of Cas12k reduces integration efficiency. • Fusion of TnsC to the C-terminus of Cas12k similarly retains high integration efficiency	
Type I-F	INTEGRATE (INsert Transposable Elements by Guide RNA-Assisted TargEting)	Comprising 3 separate plasmids: (i) pQCascade encoding tniQ-cas8/5-cas7-cas6 along with a synthetic crRNA array; (ii) pTnsABC encoding the transposase complex tnsA/tnsB/tnsC; (iii) pDonor harboring the donor DNA flanked by LE/RE.	*E. coli*	• confers >95% accuracy of DNA integration in 16 genomic sites. • accommodates genetic payloads of variable lengths up to 10 kb, although the optimal size is 775 bp.	[18]
	MUCICAT (multicopy chromosomal integration using CRISPR-associated transposases)	3 plasmids similar to the original INTEGRATE, but with different crRNA designs	*E. coli* BL21(DE3), *Tatumella citrea*	• Using a crRNA targeting the 28-copy IS1 loci, colonies that accumulate >10 copies of cargo integration at IS1 loci can be detected. • Using a crRNA array, cargo integration into 8 different genomic sites is achieved. • Achieves 12.5% of colonies harboring 8 gene copies after a round of re-streaking. The multicopy integration frequency can be increased with several rounds of re-streaking.	[50, 51]

	*Ptr*CAST-based MUCICAT	gene cassettes in the 3 plasmids are similar to those in *Vch*CAST INTEGRATE but the component proteins are derived from *Pseudoalteromonas translucida* (*Ptr*CAST) driven by different promoters such as tetracycline inducible (Tet) promoters	*E. coli* BL21Star™(DE3)	● achieves up to 100% insertion efficiency for a 15.4 kb cargo. ● achieves 12.5% of colonies harboring 8 copies after a single round of re-streaking.	[51]
	All-in-one INTEGRATE	all components of INTEGRATE are packaged in a single plasmid	*E. coli* *Pseudomonas putida*	● higher protein expression levels drives higher integration frequencies, without compromising the specificity. ● improves editing efficiency by lowering the *E. coli* culture temperature from 37°C to 30°C. ● Under the optimal conditions, a 10 kb cargo can be integrated into the genome at ≈100% efficiency.	[52]

698

Figure 1

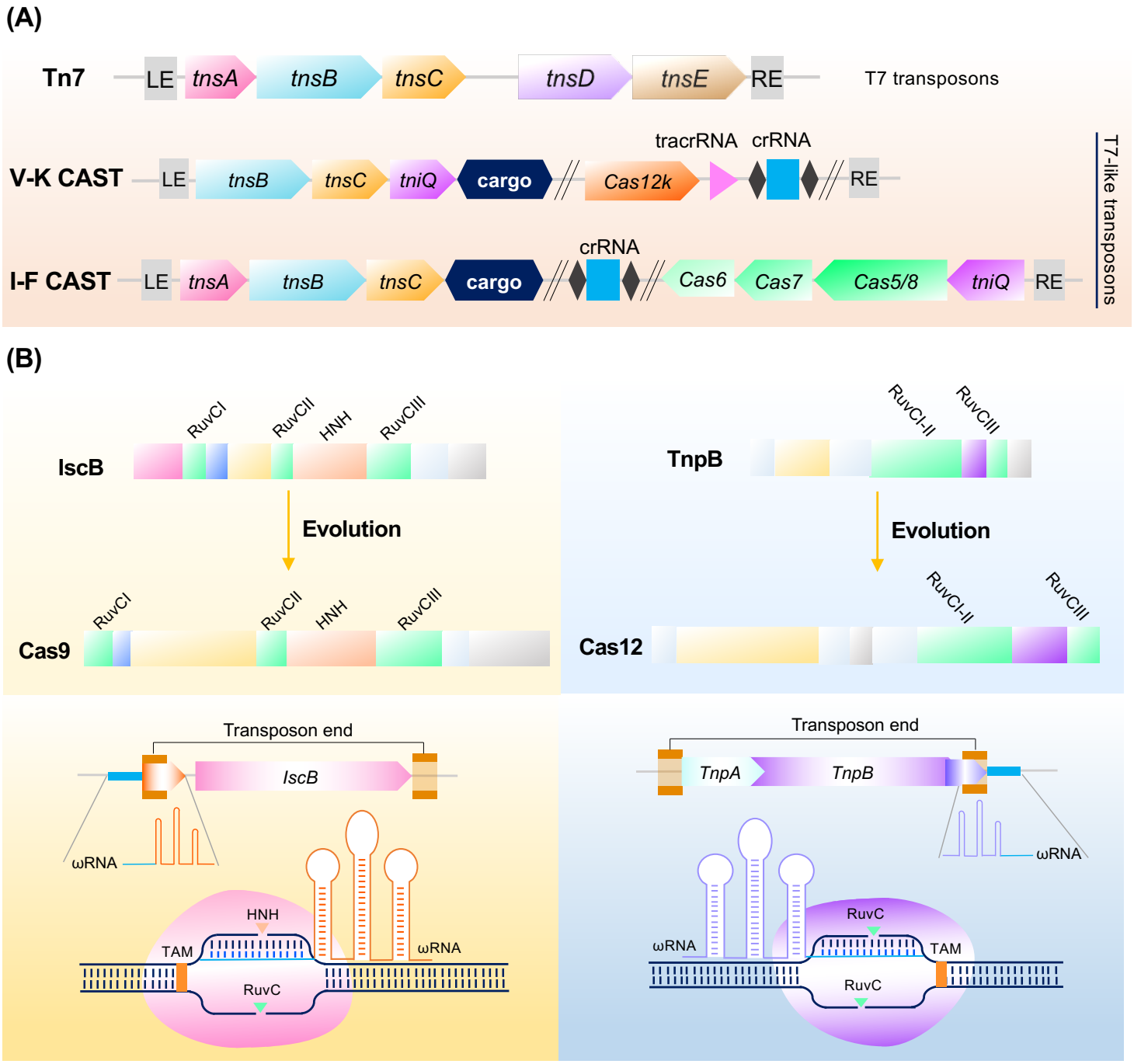

Figure 1

Figure 2

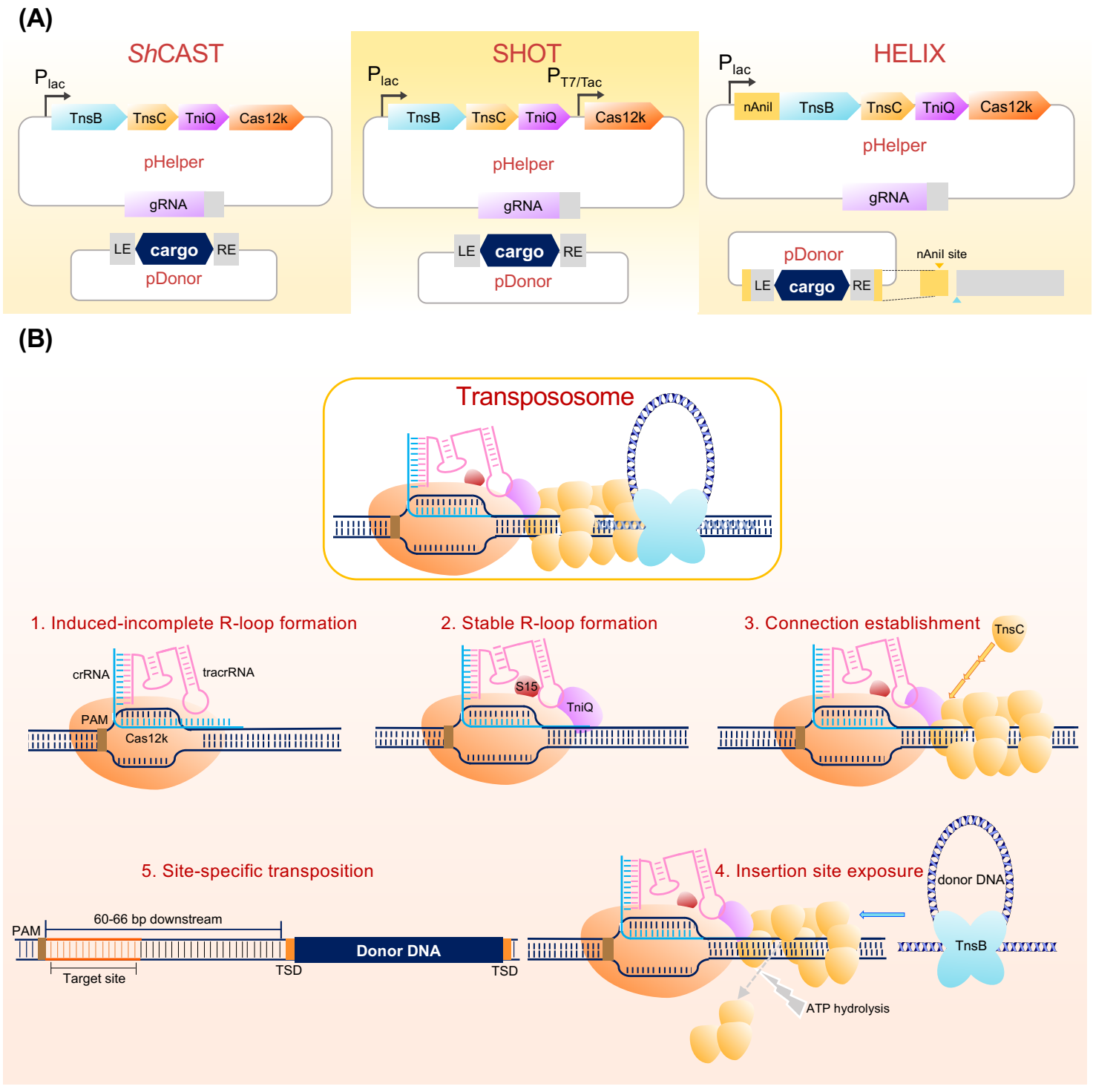

Figure 2

Figure 3

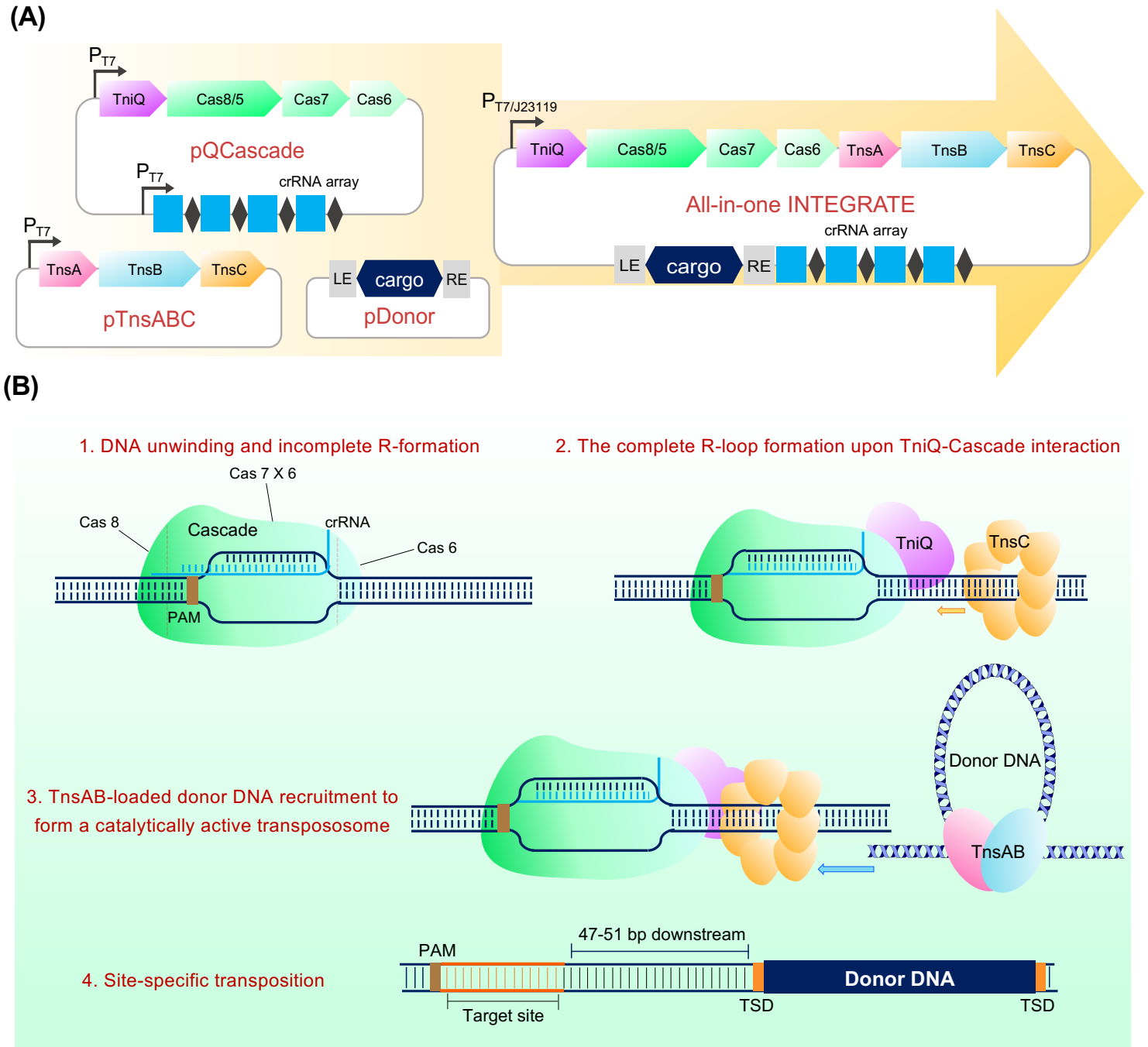

Figure 3

Figure 4

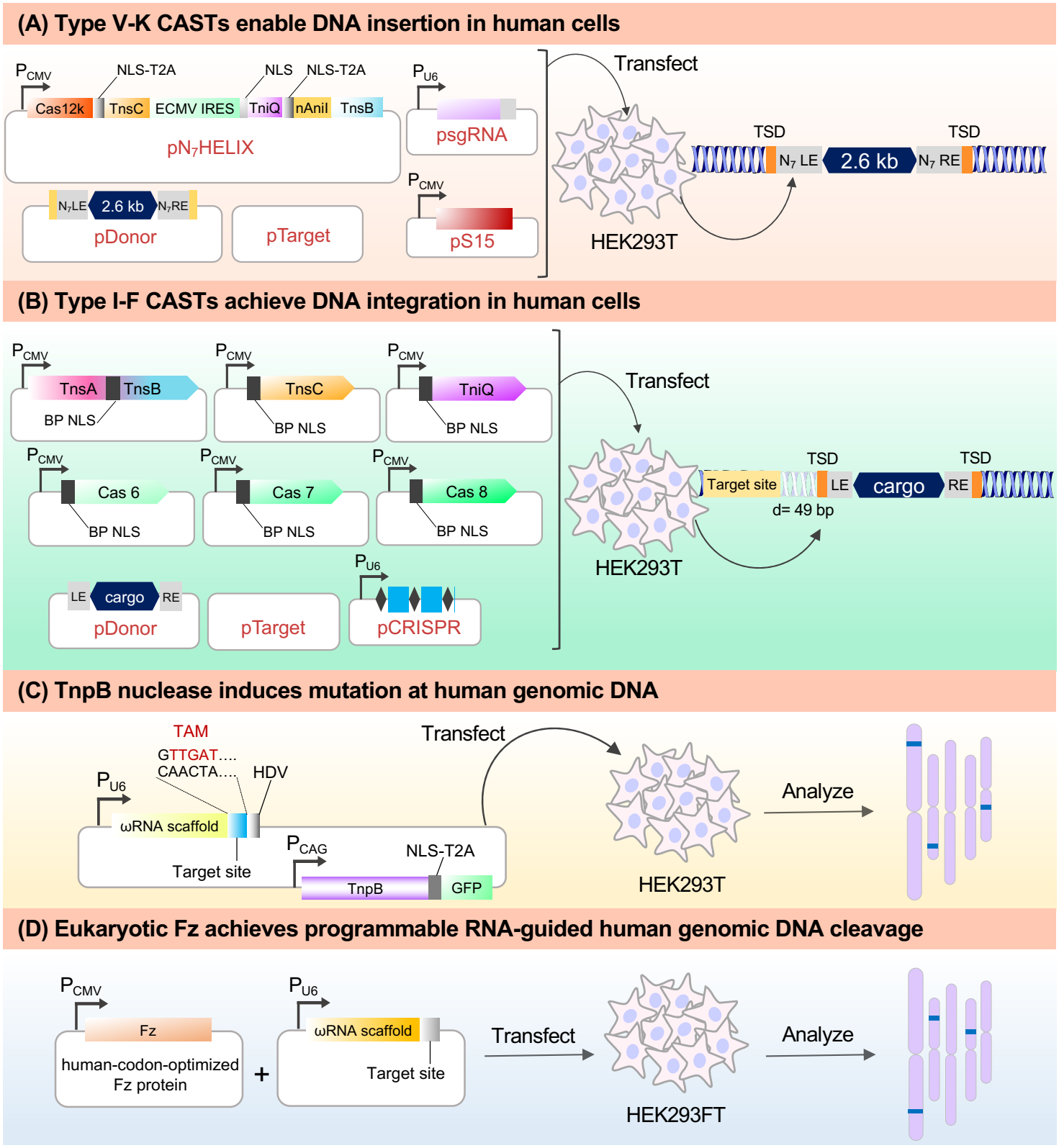

Figure 4