

1           **RNA-guided Genome Engineering: Paradigm Shift Towards Transposons**

2           Chin-Wei Chang<sup>1,ξ</sup>, Vy Anh Truong<sup>1,ξ</sup>, Nam Ngoc Pham<sup>1</sup>, and Yu-Chen Hu<sup>1,2\*</sup>, ORCID

3  
4           <sup>1</sup> Department of Chemical Engineering, National Tsing Hua University, Hsinchu, Taiwan 300

5           <sup>2</sup> Frontier Research Center on Fundamental and Applied Sciences of Matters, National Tsing Hua  
6           University, Hsinchu, Taiwan 300

7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17           **Keywords:** CASTs, OMEGA, transposon, RNA-guided genome engineering, CRISPR, double  
18           strand break

19  
20  
21  
22  
23  
24           Running Title: RNA-guided Transposons for Genome Editing

25  
26  
27  
28           <sup>ξ</sup> These authors contributed equally to this work.

29           \*Correspondence: [ychu@mx.nthu.edu.tw](mailto:ychu@mx.nthu.edu.tw) (Y.-C. Hu)

30           ORCID <https://orcid.org/0000-0002-9997-4467>

31

32 **Abstract**

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

43

## 44 **CRISPR and Transposons for Genome Engineering**

45 CRISPR-Cas systems are categorized into 2 classes, 6 types (e.g. Cas9 for type II and Cas12 for  
46 type V) and 33 subtypes [1], among which CRISPR-Cas9 is first repurposed for genome  
47 engineering of eukaryotes [2, 3] and prokaryotes [4, 5]. CRISPR-Cas9 requires **Cas9** (See  
48 Glossary), **crRNA** and **tracrRNA** (See Glossary) to recognize the **PAM** (See Glossary) sequence  
49 on the target DNA and induce double strand break (DSB), which triggers intrinsic **DNA repair**  
50 (See Glossary) and ensuing DNA insertions/deletions (indels). However, DNA repair is inefficient  
51 in non-dividing eukaryotes [6] and some bacteria [7]; DSB sparks innate immune responses in  
52 human cells [8] and may elicit genomic deletions and chromosomal translocations [9, 10].  
53 Furthermore, the large size of the most widely used Cas9 and Cas12 (1000-1300 aa) hinders their  
54 delivery by **AAV** (See Glossary) for gene therapy [11, 12]. Although miniature Cas12f homologs  
55 (400–700 aa) are found, their editing efficiencies remain low [13-15].

56 In 2017, the connection between Tn7-like **transposons** (See Glossary) and CRISPR is  
57 discovered [16], with subsequent observations revealing associations with various CRISPR types,  
58 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-  
59 associated transposons (termed CASTs) are first repurposed for DSB-free, RNA-guided DNA  
60 insertion into *E. coli*, which implicates the potentials of CASTs to obviate the requirement of DSB  
61 induction in CRISPR-Cas systems. In 2021, IscB (400-500 aa) and TnpB (408 aa) nucleases in the  
62 **insertion sequence (IS)** (See Glossary) family transposons are transformed into hypercompact  
63 programmable RNA-guided DNA nucleases [23, 24]. These nucleases, part of the OMEGA  
64 (obligate mobile element guided activity) system, show promise in overcoming the size limitations  
65 of current CRISPR-Cas nucleases.

66 Several articles have comprehensively reviewed CASTs, either offering a general introduction  
67 [25-28] or delving into specific aspects like molecular mechanisms [20, 21, 29] and applications  
68 related to DSB-free DNA integration [28, 30, 31]. Yet OMEGA system is reviewed in only two  
69 papers [21, 32] and several research spotlight articles [33-35]. This paper provides a comprehensive  
70 review of recent advances, specifically focusing on the development and applications of CASTs in  
71 genome editing for both prokaryotes and eukaryotes. Additionally, we offer an updated overview  
72 of OMEGA in the context of eukaryotic genome editing.

### 73 **What are CASTs and OMEGA**

74 Bioinformatic searches have revealed CASTs repertoire to over 1000 subsystems [36], among  
75 which types V-K and I-F are the most exhaustively studied [26]. Type V-K CASTs encode four  
76 proteins (TnsB, TnsC, TniQ, Cas12k), tracrRNA/crRNA and the genetic cargo within the left end  
77 (LE) and right end (RE) of transposon (Fig. 1A). TnsB is the transposase for integration; TnsC is  
78 an ATPase for target site selection while TniQ is an adaptor protein. Type I-F3 CASTs encode  
79 seven proteins (TnsA, TnsB, TnsC, Cas6, Cas7, Cas5/8 and TniQ) and expresses only crRNA  
80 (without tracrRNA). The cassette, along with the cargo, is embedded within the LE/RE ends (Fig.  
81 1A).

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

## 89 **CASTs for Prokaryotic Genome Engineering**

### 90 *Development of type V-K CASTs*

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

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

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

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

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

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

123 This co-integration problem is addressed in a recent study [46]. Bacterial transposons propagate  
124 through cut-and-paste or copy-and-paste pathway [47]. In the Tn7 transposon, TnsA and TnsB nick  
125 the 5' and 3' end of donor DNA, respectively, resulting in simple insertions via cut-and-paste  
126 transposition. Tou *et al.* reason that the lack of TnsA in the type V-K CAST leads to TnsB-mediated  
127 3' nicking only, thus resulting in copy-and-paste integration and undesired co-integration [46].  
128 With this hypothesis, HELIX (Table 1) is developed from *Sh*CAST by fusing a nicking  
129 endonuclease (nAnil) to the N terminus of TnsB to restore the 5' nicking capability (Fig. 2A).  
130 HELIX confers large DNA insertion into genomic targets in *E. coli* via cut-and-paste mechanism,  
131 which almost abolishes the co-integration while retaining robust integration efficiencies (Table 1).  
132 Note, however, that C-terminal fusion of nAnil to TnsB does not yield transposition. The HELIX  
133 design, derived from *Sh*CAST, is also applicable to other type V-K CASTs or a different strain of  
134 *S. hofmannii* [46].

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

### 138 ***Development of Type I CASTs***

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

147 Following the concept of INTEGRATE, MUCICAT (Table 1) systems based on different type  
148 I-F CASTs is developed for multicopy and multiplexed integration by changing the crRNA design  
149 [50, 51]. Yet, the multicopy integration does not function as effectively in *E. coli*  
150 MG1655 $\Delta$ endA $\Delta$ recA(DE3) [50]. Another type I-F CAST also fails to mediate cargo insertion  
151 [51]. These studies indicate the need to screen appropriate type I CASTs from different organisms  
152 for INTEGRATE system development.

153 To streamline INTEGRATE, Vo, *et al.* assemble all components of INTEGRATE in an all-in-  
154 one expression vector (Fig. 3A) [52]. The all-in-one INTEGRATE confers higher integration  
155 efficiencies than the original INTEGRATE, presumably because the higher efficiency of  
156 transforming a single plasmid than co-transforming 3 plasmids into single cells. DNA insertion is  
157 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.



## 178 **Applications of CASTs in Bacteria**

### 179 ***Metabolic engineering for improved product synthesis***

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

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

### 196 ***Gene editing within a complex bacterial community***

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

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

### 213 ***Targeted mutagenesis and genetic screening in diverse bacteria***

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

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

## 222 **CASTs and OMEGA for Eukaryotic Genome Engineering**

### 223 ***CASTs***

224 Tou, *et al.* discover that a type V-K CAST from *Nostoc Sp.* PCC7101 (N<sub>7</sub>CAST) enables DNA  
225 insertion in the HEK293 cell lysate [46]. The authors further construct N<sub>7</sub>HELIX (Fig. 4A), which  
226 comprises pN<sub>7</sub>HELIX to express transposition proteins, pTarget, pdonor, psgRNA, and pS15  
227 expressing S15 (a host ribosomal protein required for type V-K CASTs, Box 1). Delivering the  
228 system into HEK293T cells enables insertion of a 2.6 kb cargo into the target plasmid 57-62 bp  
229 downstream of the PAM with only 7.9% co-integrate products, a 5.3-fold improvement compared  
230 with N<sub>7</sub>CAST. However, the integration efficiency is very low (<0.1%) [46].

231 Recently Lampe, *et al.* construct pDonor, pTarget and plasmids expressing proteins and RNA  
232 components from a type I-E Cascade from *Pseudomonas sp.* S-6-2 (Fig. 4B) [60]. Delivering these  
233 plasmids into HEK293T cells enables cargo integration from pDonor into pTarget, at efficiencies  
234 <0.1%. To increase the efficiency, the authors screen 18 additional CASTs and identify a  
235 transposon from *Pseudoalteromonas sp.* (*Pse*CAST). Using the same approach and systemic  
236 engineering, the engineered *Pse*CAST confers 3-5% cargo transposition from pDonor to pTarget,  
237 with the integration efficiency decreasing with cargo size (0.8 to 15 kb). The authors further  
238 optimize *Pse*CAST and achieve donor integration into several genomic loci with efficiencies  
239 varying from 0.005-0.01%. The authors identify the essential roles of post-transposition complex  
240 (PTC) unfolding and disassembly for transposition. By co-expressing bacterial ClpX, an ATPase  
241 from MuA transposon that unfolds protein substrates, the integration efficiency is improved to ≈1%,  
242 indicating that PTC disassembly limits integration into genomic sites in HEK293T cells [60]. This  
243 work lays a foundation to exploit CASTs for eukaryotic genome engineering and demonstrates the  
244 possibility to improve the editing efficiency through systematic engineering and discovery of new

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

### 247 **OMEGA**

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

255 In the meantime, Karvelis, *et al.* uncover that TnpB from *Deinococcus radiodurans* ISDr2 (408  
256 aa) can complex with the  $\omega$ RNA ( $\approx$ 150 nt) encoded within the RE of transposon [24]. TnpB- $\omega$ RNA  
257 complex recognizes the TAM and mediates *in vitro* plasmid cleavage. Delivering the plasmid  
258 encoding TnpB and  $\omega$ RNA (Fig. 4C) into HEK293T cells induces indels at frequencies (10–20%)  
259 comparable to CRISPR–Cas9 and CRISPR-Cas12 [24].  $\omega$ RNA directs TnpB to create a staggered  
260 DNA cut in a manner similar to Cas12. TnpB represents the minimal structural and functional core  
261 of Cas12 and is the evolutionary ancestor of Cas12 [62]. TnpB can process its own mRNA into  
262  $\omega$ RNA [63], which is conserved among all guide RNAs of Cas12 [64].

263 Recently, Xiang *et al.* identify new TnpB proteins, particularly ISAam1 (369 aa) and ISYmu1  
264 (382 aa), and showcase their ability to induce indels across genomic loci in HEK293T cells [65].  
265 ISAam1 and ISYmu1 outperform 3 Cas12f editors, exhibiting editing efficiency comparable to  
266 SaCas9 (1,053 aa) but with a significantly smaller size. They predominantly induce deletions, and  
267 insertions are less common, in a way similar to those in CRISPR-Cas systems. Notably, the

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

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

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

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

### 297 **Concluding Remarks and Future Perspectives**

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

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

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

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

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

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

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

344



## Glossary

345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366

**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  $\approx 4.7$  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 have different PAM preferences.

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

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

379

380

## Box

381 **Box 1. Mechanisms of type V-K CASTs transposition:** In type V-K CASTs, the overall structure  
382 of Cas12k resembles other Cas12 proteins. Cryo-electron microscopy (cryo-EM) studies show that  
383 Cas12k adopts a bi-lobed structure connected by a loop [80, 81]. The N-terminal lobe is composed  
384 of the wedge (WED), recognition (REC) and PAM interacting (PI) domains, in which WED  
385 domain is crucial for recognizing crRNA/tracrRNA. The tracrRNA serves as a scaffold to position  
386 Cas12k and crRNA for pinpointing complementary PAM sequences [80, 81]. Upon binding to a  
387 5'-GTN-3' PAM, Cas12k initiates RNA hybridization to the target strand DNA, hence inducing  
388 incomplete R-loop formation [82]. TniQ recognizes tracrRNA and R-loop to bridge tracrRNA and  
389 TnsC filament [82].

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

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

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

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

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

424 **Box 2. Mechanisms of type I CASTs transposition:** In type I-F3 *Vch*CAST, one Cas6 subunit,  
425 six subunits of Cas7 and one Cas8/5 molecule form the Cascade complex and associate with crRNA  
426 [18, 88]. Cascade binds DNA and TniQ, recruits TnsC and the heteromeric TnsAB transposase in  
427 complex with the donor DNA, to assemble the catalytically active transpososome (Fig. 3B) [87].  
428 The Cas8/5 fusion protein binds the 5' crRNA handle and contacts the TniQ dimer [18]. The  
429 dimeric TniQ binds to the Cascade complex in a head-to-tail configuration at the interface formed  
430 by Cas6 and Cas7 near the 3' end of crRNA [89]. One TniQ interacts with Cas6 whereas the other  
431 TniQ interacts with Cas7 [88]. The Cascade-TniQ complex results in the complete R-loop  
432 formation and scans numerous sites in the target DNA to specify the site of DNA integration [20].  
433 Notably, TniQ-Cascade complex alone binds not only at the target site specified by Cascade's  
434 crRNA, but also many off-target sites with mismatches [90]. Most off-target sites fail to recruit

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

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

444 Similar to CRISPR-Cas systems, the crRNA/tracrRNA or crRNA in CASTs can be engineered  
445 as the chimeric gRNA carrying the guide sequence for PAM recognition [92]. CASTs systems are  
446 more complex than the class 2 CRISPR-Cas systems, but do not require DSB induction. Both type  
447 V-K and I-F CASTs encode a core transposition machinery composed of a transposase TnsB for  
448 integration, a AAA+ ATPase TnsC for target site selection, and an adaptor protein TniQ [17, 18,  
449 22, 78]. Despite the similarities, differences between these two CASTs exist [21, 26, 30]. Type V-  
450 K TnsC forms continuous helical filaments on DNA [84], whereas type I-F3 TnsC forms  
451 heptameric structure [90]. Moreover, type V-K CASTs require host ribosomal protein S15 while  
452 type I CASTs require a host-encoded integration host factor (IHF).

453

## 454 **Figure Captions**

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

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

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

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

481 **Figure 4.** CASTs and OMEGA for genome engineering in eukaryotes. (A) Type V-K CAST.  
482 The N<sub>7</sub>HELIX enables insertion of a 2.6 kb into the target plasmid 57-62 bp downstream of the  
483 PAM in the HEK293T cells. The proteins are fused with bipartite (BP) NLS and may be linked by  
484 2A peptides. (B) Type I CAST. Co-transfection of pDonor harboring the mini-transposon, pTarget  
485 and plasmids expressing proteins and RNA components into HEK293T cells enables cargo  
486 integration from the pDonor plasmid into pTarget. (C) OMEGA-encoded TnpB for genome  
487 engineering. The plasmid encoding TnpB and ωRNA optimized for expression in eukaryotic cells  
488 is transfected into HEK293T cells and induces mutation. (D) Fn-mediated genome editing in  
489 HEK293FT cells. Transfection of plasmids encoding Fz and ωRNA induces RNA-guided DNA  
490 cleavage in the genome of human cells.

491

492

### 493 **Competing interests**

494 The authors declare no competing interests.

### 495 **Author contributions**

496 Chang, C.-W., Pham, N.M. and Truong, A.V draw the figures, prepare tables and write the paper.

497 Hu, Y.-C. supervises the project and writes the paper.

### 498 ***Acknowledgement***

499 The authors acknowledge the financial support from the National Science and Technology  
500 Council (NSTC 112-2223-E-007-002, 112-2314-B-007-004-MY3, 111-2223-E-007-003, 111-  
501 2634-F-007-007, 111-2622-E-007-003, 111-2622-8-007-011) and National Health Research  
502 Institutes (NHRI-EX112-11014BI, NHRI-EX113-11329EI), Taiwan.

503 **References**

- 504 1. Makarova, K.S. *et al.* (2020) Evolutionary classification of CRISPR–Cas systems: a burst of  
505 class 2 and derived variants. *Nat. Rev. Microbiol.* 18, 67-83.
- 506 2. Cong, L. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* 339,  
507 819-23.
- 508 3. Mali, P. *et al.* (2013) RNA-guided human genome engineering via Cas9. *Science* 339, 823-826.
- 509 4. Jinek, M. *et al.* (2012) A programmable dual-RNA–guided DNA endonuclease in adaptive  
510 bacterial immunity. *Science* 337, 816-821.
- 511 5. Jiang, W., Bikard, D., Cox, D., Zhang, F. and Marraffini, L.A. (2013) RNA-guided editing of  
512 bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* 31, 233-239.
- 513 6. Arnoult, N. *et al.* (2017) Regulation of DNA repair pathway choice in S and G2 phases by the  
514 NHEJ inhibitor CYREN. *Nature* 549, 548-552.
- 515 7. Chung, M.E. *et al.* (2017) Enhanced integration of large DNA into *E. coli* chromosome by  
516 CRISPR/Cas9. *Biotechnol. Bioeng.* 114, 172-183.
- 517 8. Kim, S. *et al.* (2018) CRISPR RNAs trigger innate immune responses in human cells. *Genome*  
518 *Res.* 28, 367-373.
- 519 9. Adikusuma, F. *et al.* (2018) Large deletions induced by Cas9 cleavage. *Nature* 560, E8-E9.
- 520 10. Kosicki, M., Tomberg, K. and Bradley, A. (2018) Repair of double-strand breaks induced by  
521 CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat Biotechnol* 36, 765-  
522 771.
- 523 11. Doudna, J.A. (2020) The promise and challenge of therapeutic genome editing. *Nature* 578,  
524 229-236.
- 525 12. Rousset, F. and Sorek, R. (2021) A treasure trove of molecular scissors. *Science* 374, 37-38.



- 526 13. Kim, D.Y. *et al.* (2022) Efficient CRISPR editing with a hypercompact Cas12f1 and engineered  
527 guide RNAs delivered by adeno-associated virus. *Nat. Biotechnol.* 40, 94-102.
- 528 14. Wu, Z. *et al.* (2021) Programmed genome editing by a miniature CRISPR-Cas12f nuclease.  
529 *Nat Chem Biol* 17, 1132-1138.
- 530 15. Xu, X. *et al.* (2021) Engineered miniature CRISPR-Cas system for mammalian genome  
531 regulation and editing. *Mol. Cell* 81, 4333-4345.e4.
- 532 16. Peters, J.E., Makarova, K.S., Shmakov, S. and Koonin, E.V. (2017) Recruitment of CRISPR-  
533 Cas systems by Tn7-like transposons. *Proc. Nat. Acad. Sci.* 114, E7358-E7366.
- 534 17. Strecker, J. *et al.* (2019) RNA-guided DNA insertion with CRISPR-associated transposases.  
535 *Science* 365, 48-53.
- 536 18. Klompe, S.E., Vo, P.L.H., Halpin-Healy, T.S. and Sternberg, S.H. (2019) Transposon-encoded  
537 CRISPR–Cas systems direct RNA-guided DNA integration. *Nature* 571, 219-225.
- 538 19. Klompe, S.E. *et al.* (2022) Evolutionary and mechanistic diversity of Type I-F CRISPR-  
539 associated transposons. *Mol. Cell* 82, 616-628.e5.
- 540 20. Koonin, E.V. and Krupovic, M. (2023) New faces of prokaryotic mobile genetic elements:  
541 Guide RNAs link transposition with host defense mechanisms. *Curr. Opin. Syst. Biol.* 36,  
542 100473.
- 543 21. Tenjo-Castaño, F., Montoya, G. and Carabias, A. (2022) Transposons and CRISPR: Rewiring  
544 gene editing. *Biochemistry*.
- 545 22. Hsieh, S.C. and Peters, J.E. (2023) Discovery and characterization of novel type I-D CRISPR-  
546 guided transposons identified among diverse Tn7-like elements in cyanobacteria. *Nucleic*  
547 *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  $\omega$ RNA 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., Mougias, 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.

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

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

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

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

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

625 59. Rodriguez, L.T., Ellington, A.J., Reisch, C.R. and Chevrette, M.G. (2023) CRISPR-associated  
626 transposase for targeted mutagenesis in diverse proteobacteria. *ACS Synth. Biol.* 12, 1989-  
627 2003.

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

630 61. Kato, K. *et al.* (2022) Structure of the IscB- $\omega$ RNA ribonucleoprotein complex, the likely  
631 ancestor of CRISPR-Cas9. *Nat Commun* 13, 6719.

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

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

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

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

640 66. Bao, W. and Jurka, J. (2013) Homologues of bacterial TnpB\_IS605 are widespread in diverse  
641 eukaryotic transposable elements. Mobile DNA 4, 12.

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

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

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

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

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

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

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

656 74. Hsu, M.-N. *et al.* (2019) CRISPR technologies for stem cell engineering and regenerative  
657 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. Spectr.* 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 <i>Sh</i> 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.	<i>E. coli</i> Pir	<ul style="list-style-type: none"> <li>● integrates cargo into <math>\approx 60\%</math> of selected genomic sites with efficiencies varying from <math>&lt;10\%</math> to <math>\approx 80\%</math></li> <li>● integrates a cargo up to 10 kb.</li> </ul>	[17]
	Improved <i>Sh</i> CAST	The <i>lac</i> promoter in the pHelper plasmid is swapped with promoters of different strengths to drive the tnsB-tnsC-tniQ-Cas12k-gRNA expression.	<i>Shewanella oneidensis</i> MR-1	<ul style="list-style-type: none"> <li>● inserts a 2 kb cargo into 4 identical repeats at <math>\approx 100\%</math> efficiency.</li> <li>● enables insertion of DNA up to 31.7 kb into genome at <math>\approx 100\%</math> efficiency in a single orientation</li> </ul>	[43]
	SHOT ( <i>Sh</i> CAST-based Optimized Transposon)	<ul style="list-style-type: none"> <li>● increasing Cas12k expression, independent of TnsC/TnsB/TniQ, under a stronger promoter such as T7 or Tac.</li> <li>● Decreasing the editing temperature from 37°C to 30°C and re-streaking the colonies.</li> </ul>	<i>E. coli</i> BL21(DE3), MG1655, W3110, W, BW25113 strains and <i>Pseudomonas putida</i>	<ul style="list-style-type: none"> <li>● enhances on-target integration efficiency to <math>&gt;90\%</math> in different bacteria strains.</li> <li>● alleviates off-target effects for cargo gene into 3 tested loci and enables on-target integration of 14.5 DNA cargo with <math>\approx 100\%</math> efficiency</li> </ul>	[42]
	C12KGET (Cas12k-based genetic engineering toolkit)	stronger J23119 promoter to improve the expression of transposase and Cas12k.	<i>Sinorhizobium meliloti</i>	<ul style="list-style-type: none"> <li>● achieves genomic integration of fragments up to 10 kb in size with up to 100% efficiency</li> </ul>	[45]
	HELIX (Homing Endonuclease-assisted Large-sequence Integrating	fusing nAnil to the N terminus of TnsB.	<i>E. coli</i> pir1, <i>Scytonema hofmannii</i>	<ul style="list-style-type: none"> <li>● enables cut-and-paste DNA insertion with up to 99.4% simple insertion product purity, while retaining robust integration efficiencies on genomic targets.</li> <li>● enables integration of DNA at least 9.8 kb into target genomic site with comparably high efficiency.</li> </ul>	[46]

	CAST complexX)			<ul style="list-style-type: none"> <li>● C-terminal fusions of one or two copies of TniQ to Cas12k retains up to 62-69% integration efficiency relative to that of unfused <i>Sh</i>CAST, while fusion of either TniQ or TnsC to the N-terminus of Cas12k reduces integration efficiency.</li> <li>● Fusion of TnsC to the C-terminus of Cas12k similarly retains high integration efficiency</li> </ul>	
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.	<i>E. coli</i>	<ul style="list-style-type: none"> <li>● confers &gt;95% accuracy of DNA integration in 16 genomic sites.</li> <li>● accommodates genetic payloads of variable lengths up to 10 kb, although the optimal size is 775 bp.</li> </ul>	[18]
	MUCICAT (multicopy chromosomal integration using CRISPR-associated transposases)	3 plasmids similar to the original INTEGRATE, but with different crRNA designs	<i>E. coli</i> BL21(DE3), <i>Tatumella citrea</i>	<ul style="list-style-type: none"> <li>● Using a crRNA targeting the 28-copy IS1 loci, colonies that accumulate &gt;10 copies of cargo integration at IS1 loci can be detected.</li> <li>● Using a crRNA array, cargo integration into 8 different genomic sites is achieved.</li> <li>● 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.</li> </ul>	[50, 51]

	<i>Ptr</i> CAST-based MUCICAT	gene cassettes in the 3 plasmids are similar to those in <i>Vch</i> CAST INTEGRATE but the component proteins are derived from <i>Pseudoalteromonas translucida</i> ( <i>Ptr</i> CAST) driven by different promoters such as tetracycline inducible (Tet) promoters	<i>E. coli</i> BL21Star™(DE3)	<ul style="list-style-type: none"> <li>● achieves up to 100% insertion efficiency for a 15.4 kb cargo.</li> <li>● achieves 12.5% of colonies harboring 8 copies after a single round of re-streaking.</li> </ul>	[51]
	All-in-one INTEGRATE	all components of INTEGRATE are packaged in a single plasmid	<i>E. coli Pseudomonas putida</i>	<ul style="list-style-type: none"> <li>● higher protein expression levels drives higher integration frequencies, without compromising the specificity.</li> <li>● improves editing efficiency by lowering the <i>E. coli</i> culture temperature from 37°C to 30°C.</li> <li>● Under the optimal conditions, a 10 kb cargo can be integrated into the genome at ≈100% efficiency.</li> </ul>	[52]

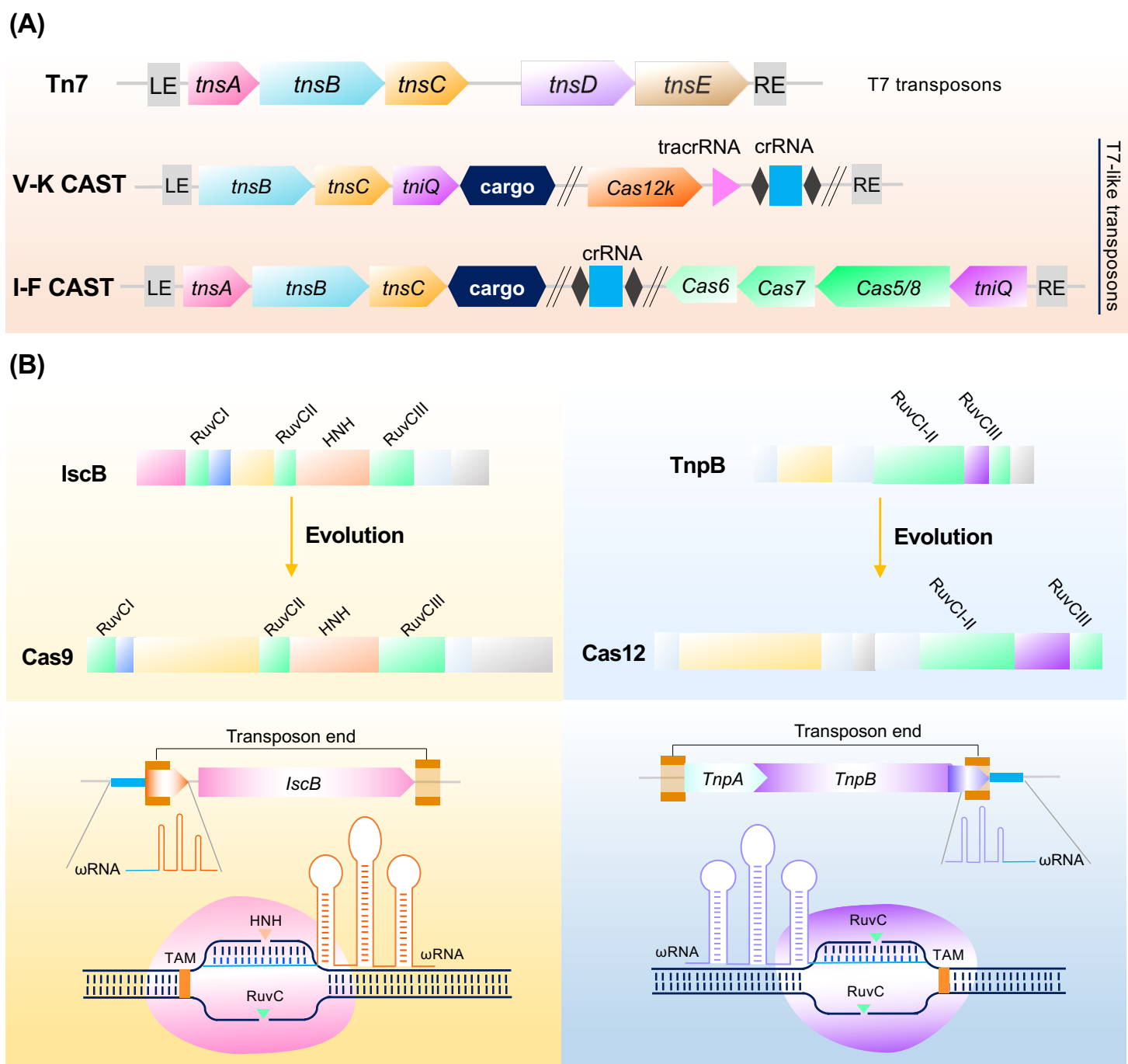


Figure 1

Figure 2

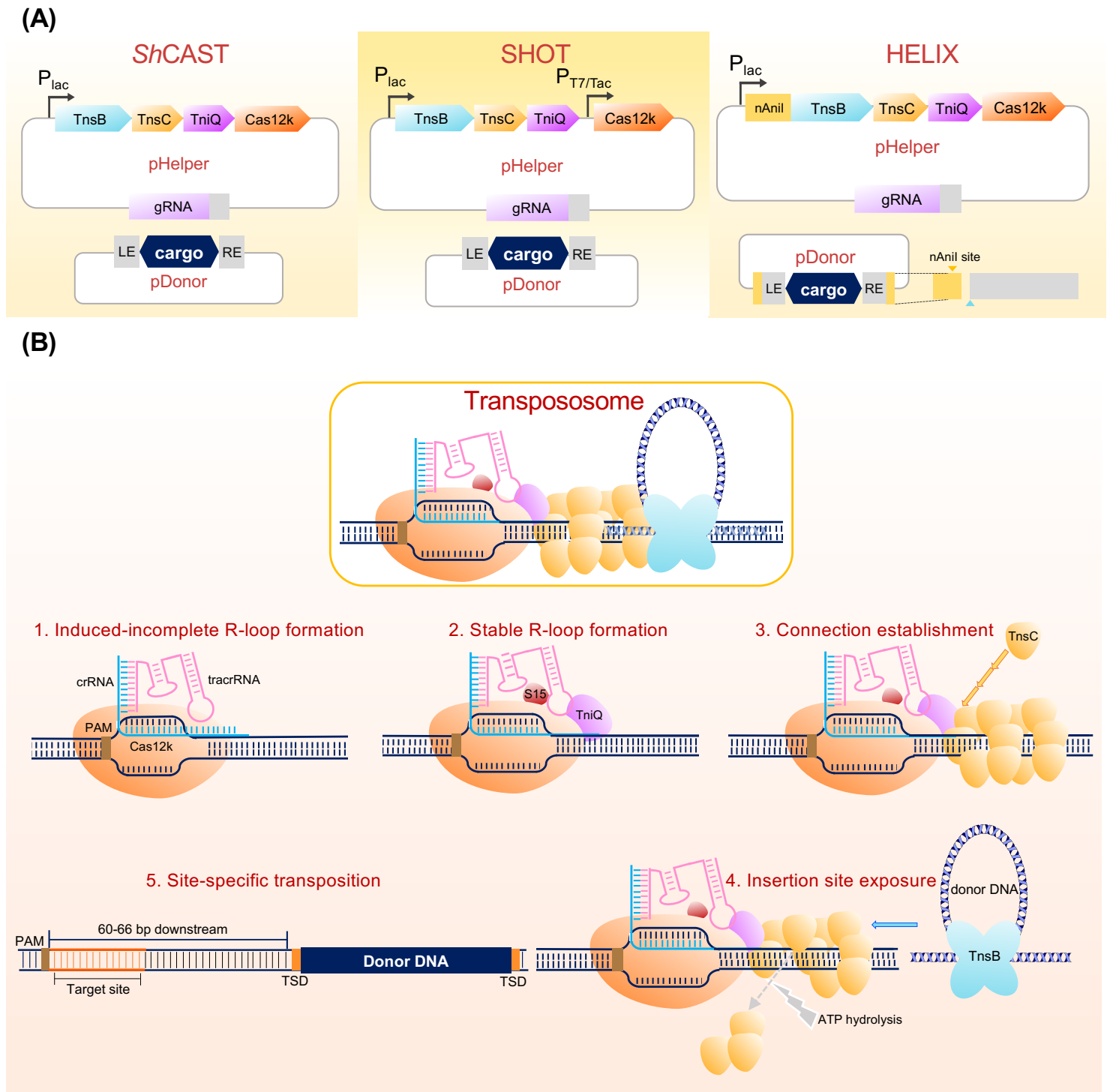


Figure 2

Figure 3

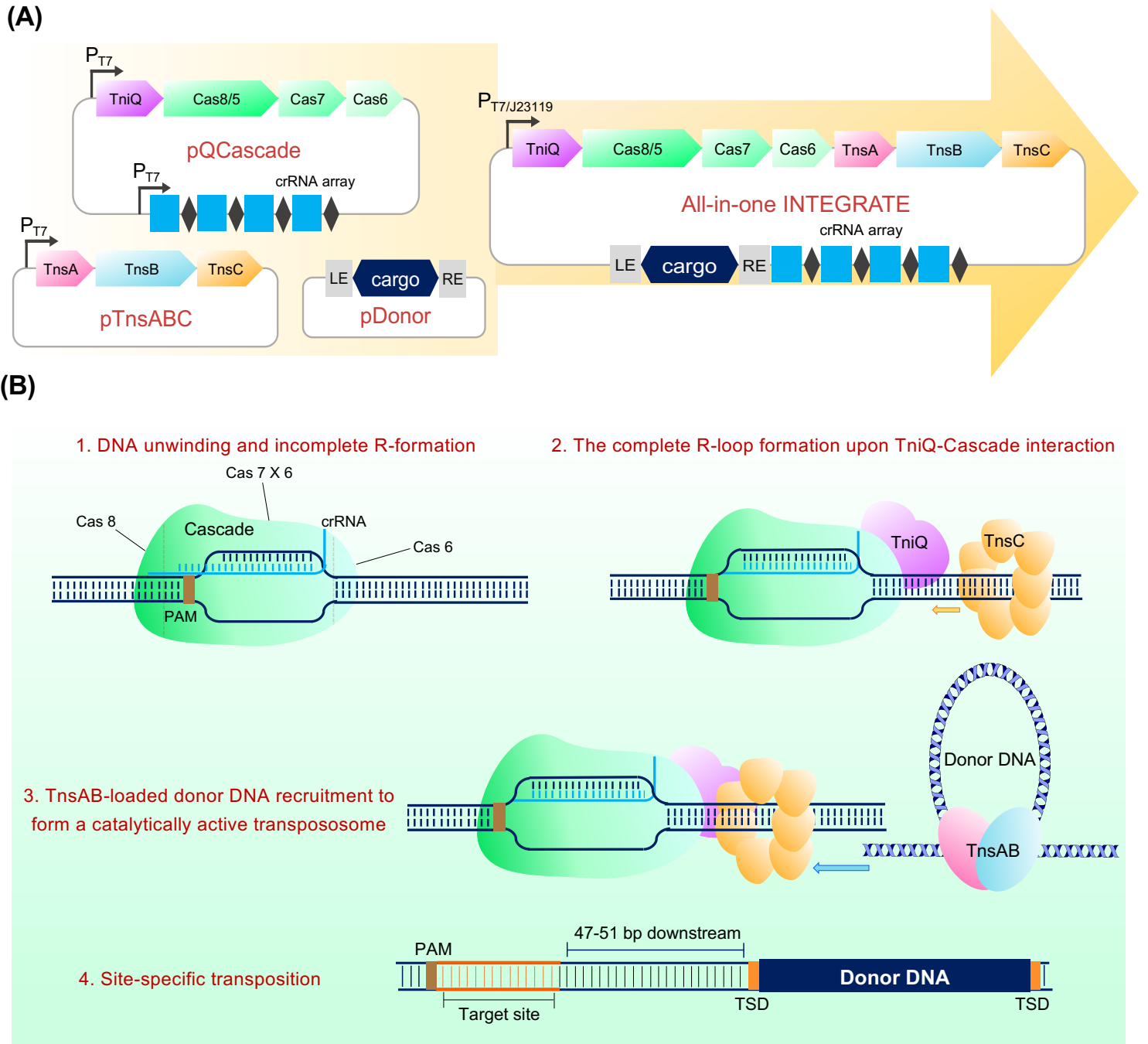


Figure 3

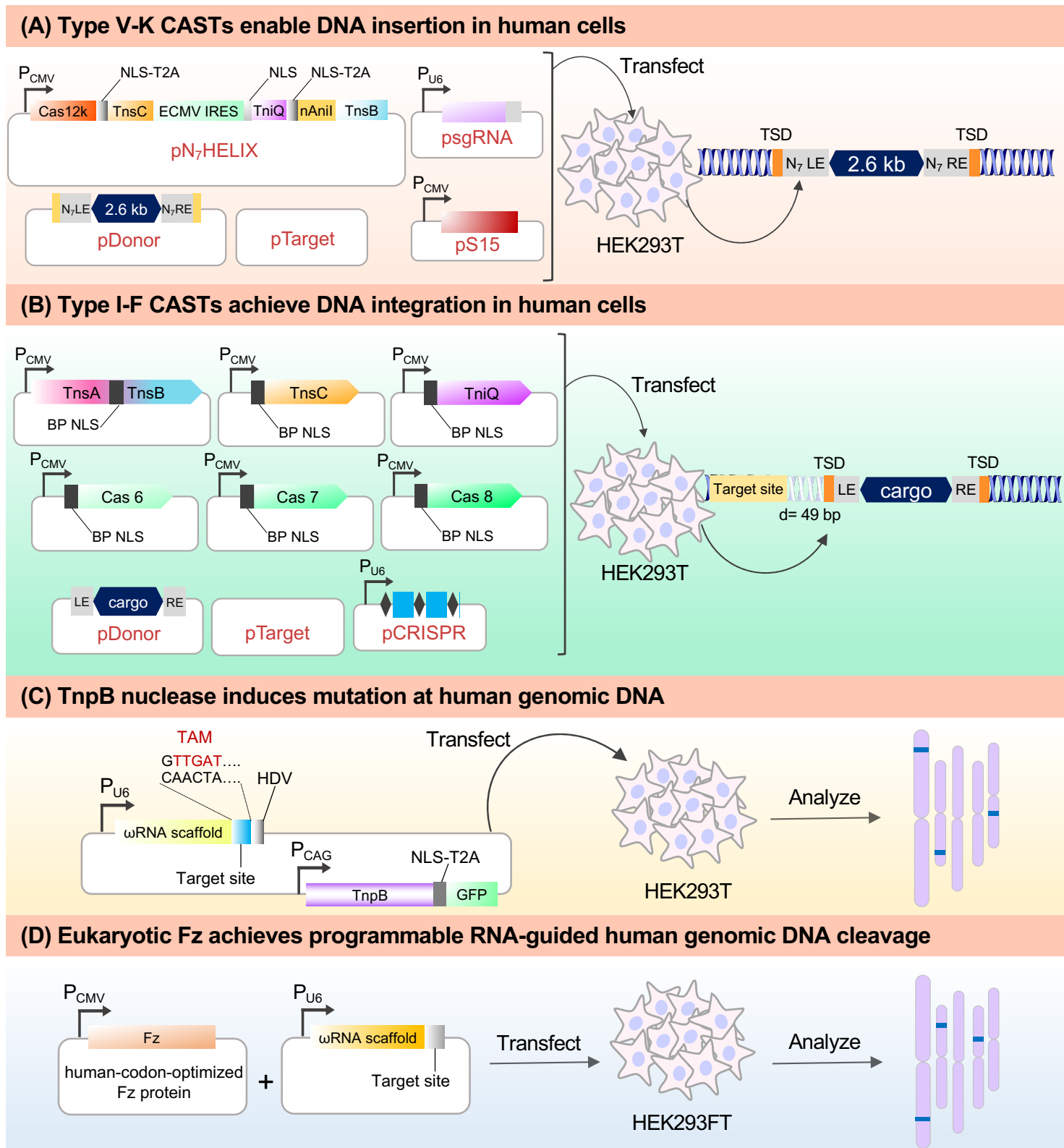


Figure 4