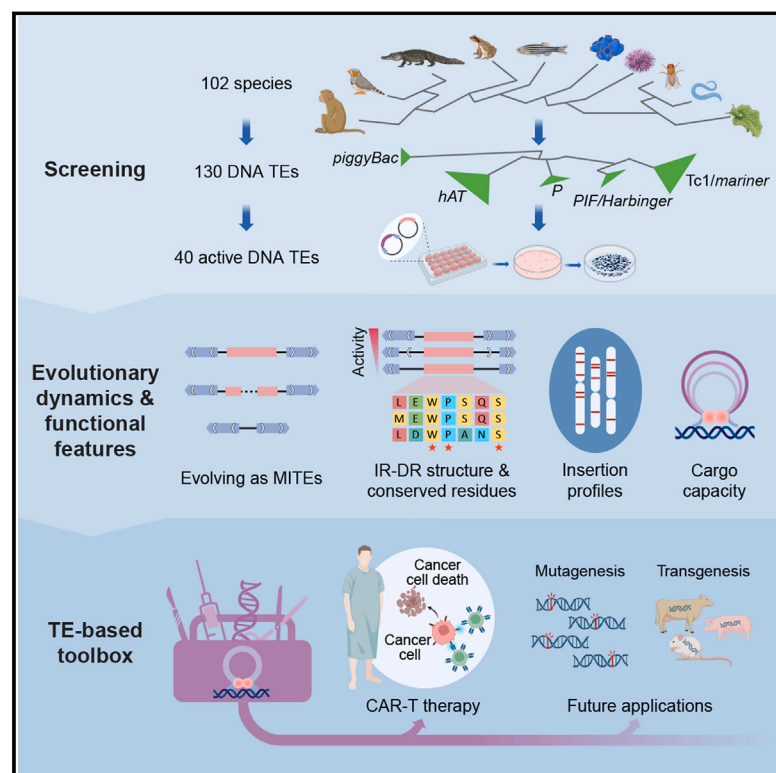


Heterologous survey of 130 DNA transposons in human cells highlights their functional divergence and expands the genome engineering toolbox

Graphical abstract



Authors

Tongtong Zhang, Shengjun Tan, Na Tang, ..., Baocheng Guo, Haoyi Wang, Yong E. Zhang

Correspondence

wanghaoyi@ioz.ac.cn (H.W.), zhangyong@ioz.ac.cn (Y.E.Z.)

In brief

A heterologous survey of 130 DNA transposable elements (TEs) in human cells not only unveils evolutionary and functional features linked to transposition activity but also expands the TE-based genome engineering toolbox by incorporating functionally diverse TEs, particularly the highly potent MAG.

Highlights

- Forty active DNA TEs from mining 102 metazoan genomes increase the TE toolbox diversity
- *Tc1* elements with high copy numbers and IR-DR structures tend to show high activity
- Active TEs show diverged functional features such as insertion profile and cargo capacity
- MAG outperforms lentiviral vectors in CAR-T cell therapy



Article

Heterologous survey of 130 DNA transposons in human cells highlights their functional divergence and expands the genome engineering toolbox

Tongtong Zhang,^{1,2,3,12} Shengjun Tan,^{4,12} Na Tang,^{1,3,5,12} Yuanqing Li,^{2,4,12} Chenze Zhang,⁶ Jing Sun,^{1,2,3} Yanyan Guo,^{2,4} Hui Gao,⁷ Yujia Cai,^{1,2,3} Wen Sun,^{1,3,5} Chenxin Wang,^{1,3,5} Liangzheng Fu,^{1,2,3} Huijing Ma,⁴ Yachao Wu,^{1,2,3} Xiaoxuan Hu,^{1,2,3} Xuechun Zhang,^{1,2,3} Peter Gee,⁸ Weihua Yan,⁹ Yahui Zhao,⁴ Qiang Chen,¹⁰ Baocheng Guo,^{2,4,11} Haoyi Wang,^{1,2,3,5,*} and Yong E. Zhang^{2,4,13,*}

¹Key Laboratory of Organ Regeneration and Reconstruction, State Key Laboratory of Stem Cell and Reproductive Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China

²University of Chinese Academy of Sciences, Beijing 100049, China

³Institute for Stem Cell and Regeneration, Chinese Academy of Sciences, Beijing 100101, China

⁴Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China

⁵Beijing Institute for Stem Cell and Regenerative Medicine, Beijing 100101, China

⁶National Key Laboratory of Efficacy and Mechanism on Chinese Medicine for Metabolic Diseases, Beijing University of Chinese Medicine, Beijing 100029, China

⁷Rengene Biotechnology Co., Ltd., Beijing 100036, China

⁸MaxCyte Inc., Rockville, MD 20850, USA

⁹Cold Spring Biotech Corp., Beijing 100031, China

¹⁰Department of Biotherapy, Cancer Center and State Key Laboratory of Biotherapy, West China Hospital, Sichuan University, Chengdu 610041, China

¹¹Academy of Plateau Science and Sustainability, Qinghai Normal University, Xining 810008, China

¹²These authors contributed equally

¹³Lead contact

*Correspondence: wanghaoyi@ioz.ac.cn (H.W.), zhangyong@ioz.ac.cn (Y.E.Z.)

<https://doi.org/10.1016/j.cell.2024.05.007>

SUMMARY

Experimental studies on DNA transposable elements (TEs) have been limited in scale, leading to a lack of understanding of the factors influencing transposition activity, evolutionary dynamics, and application potential as genome engineering tools. We predicted 130 active DNA TEs from 102 metazoan genomes and evaluated their activity in human cells. We identified 40 active (integration-competent) TEs, surpassing the cumulative number (20) of TEs found previously. With this unified comparative data, we found that the Tc1/*mariner* superfamily exhibits elevated activity, potentially explaining their pervasive horizontal transfers. Further functional characterization of TEs revealed additional divergence in features such as insertion bias. Remarkably, in CAR-T therapy for hematological and solid tumors, *Mariner2_AG* (MAG), the most active DNA TE identified, largely outperformed two widely used vectors, the lentiviral vector and the TE-based vector SB100X. Overall, this study highlights the varied transposition features and evolutionary dynamics of DNA TEs and increases the TE toolbox diversity.

INTRODUCTION

In the 1940s, Barbara McClintock discovered the first transposable element (TE), consisting of a pair of DNA transposons harboring terminal inverted repeat (TIR), autonomous *Ac*, and non-autonomous *Ds*.¹ McClintock found that TEs act as “controlling elements” by affecting the expression of neighboring genes and as mutagens by inducing changes in host genomes.^{2,3} Over the subsequent 70 years, the field experienced an expansion, with TEs discovered to inhabit nearly all eukaryotic genomes.^{4–6} Among various types of TEs, DNA

TEs include numerous superfamilies, such as *hAT* with *Ac/Ds* as one founding member and Tc1/*mariner*. The contribution of DNA TEs to their hosts has attracted broad interests where they are domesticated in various important pathways.^{7–10}

Meanwhile, DNA TEs have been studied from three other angles, including factors underlying transposition activity, evolutionary dynamics, and genome engineering tool development. Despite the limited scale of these studies, they have contributed essential knowledge. First, coding and noncoding determinants of transposition have been investigated. Residues including D (aspartic acid), D, and E/D (E for glutamic acid) have been identified as the catalytic



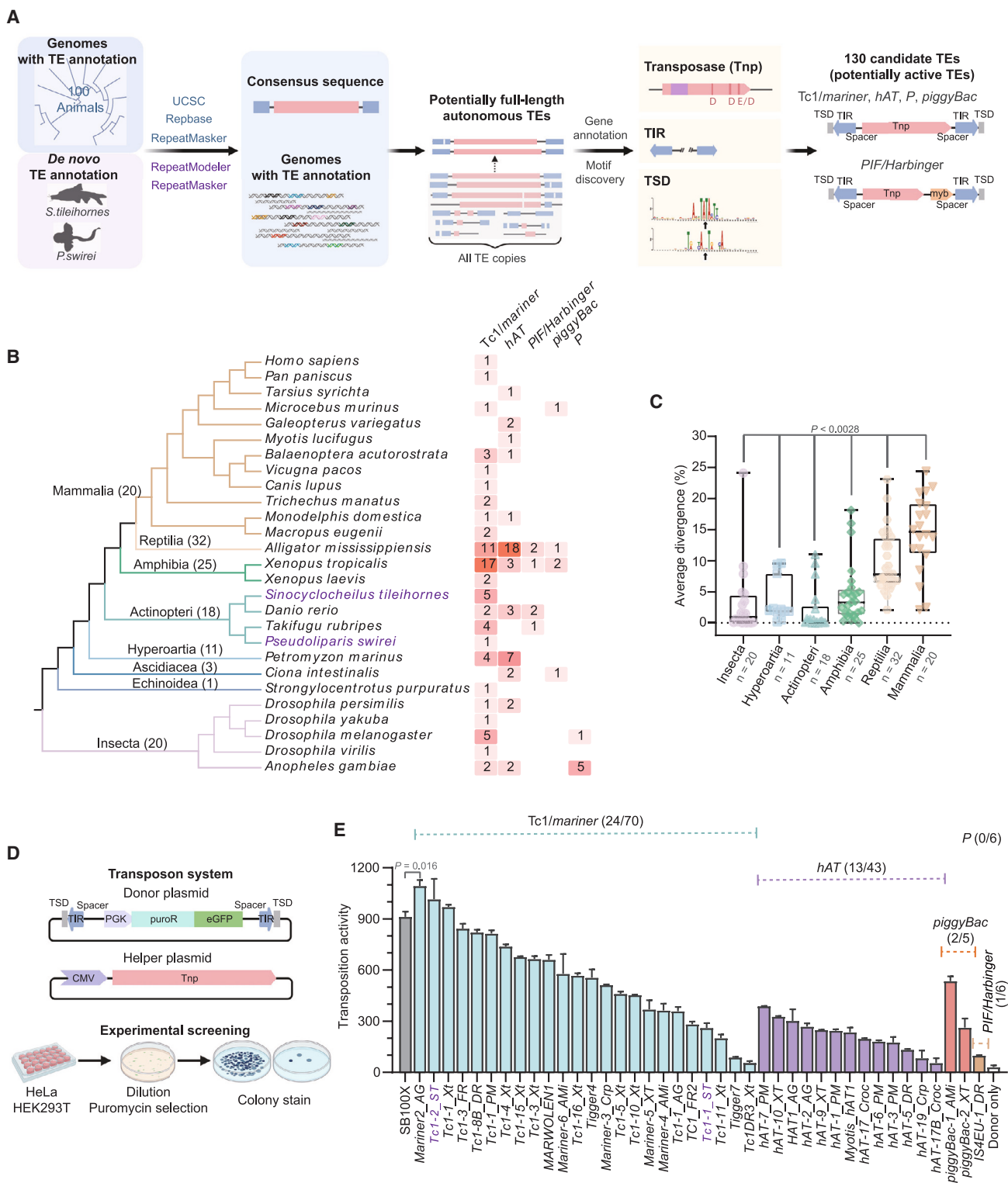


Figure 1. Mining of active TEs in 102 animal genomes and functional screening

(A) The pipelines for predicting active TEs and annotating coding and noncoding sequences.

(B) The distribution of 130 potentially active DNA TEs across species and superfamilies. The count of TEs belonging to each animal class is listed in parentheses. The numbers in each superfamily are color-coded.

(legend continued on next page)

core.^{11–13} TIRs are bound by transposases,^{14–16} and some Tc1/*mariner* elements show characteristic structure with two imperfect direct repeats (DRs) as the binding sites within each TIR (IR-DR).^{11,17–19} Second, DNA TEs are believed to undergo “horizontal transfer, vertical inactivation, and stochastic loss.”^{20–22} That is, an autonomous TE invades a host followed by repression and the accumulation of inactivating mutations.^{4,21,23–25} Furthermore, non-autonomous TEs including miniature inverted repeat TEs (MITEs, 50–800 bp internally deleted non-autonomous TEs) efficiently compete for transposases encoded by related autonomous TEs due to the presence of TIRs and their small size.^{26–29} Third, multiple DNA TEs have been developed as tools for versatile applications, including insertional mutagenesis and transgenesis.^{30–35} Among these TEs, *Sleeping Beauty* (SB) from the Tc1/*mariner* superfamily and *piggyBac* (PB) from the *piggyBac* superfamily have drawn most attention.^{36,37} With extensive optimization of coding and/or noncoding sequences, active variants (e.g., SB100X) have been developed as potent non-viral vectors.^{32,38} SB100X-based chimeric antigen receptor (CAR) T cell therapy has been employed to treat hematological tumors.^{39–41}

Clearly, the bias toward a small number of DNA TEs and heterogeneity across experimental studies make it difficult to identify general rules across TE superfamilies or families. Specifically, the transposition activity of DNA TEs after their heterologous expression in human cells remains unpredictable. How evolutionary features (e.g., superfamily type) and sequence features (e.g., IR-DR) affect transposition are largely unknown. Regarding evolutionary dynamics, while the amplification of MITEs as deletion derivatives of autonomous TEs has been reported,^{26,27,42} whether TE superfamilies are generally associated with MITEs is less clear. From an application standpoint, head-to-head comparisons of DNA TEs concerning cargo size tolerance or integration patterns are lacking.⁴³ Furthermore, TE-based genome engineering tools with diverse functional characteristics, particularly highly potent CAR-T vectors effectively treating both hematological and solid tumors, remain underdeveloped.

To fill these gaps, we performed large-scale experiments of DNA TEs in human cells to achieve two goals: (1) to derive general rules about factors underlying transposition activity and evolutionary dynamics and (2) to expand the TE toolbox. Specifically, by mining 102 metazoan genomes, we predicted 130 putatively active DNA TEs spanning five superfamilies. Through functional screening in human cells, we identified 40 active TEs, twice the total number (20) of previously reported active DNA TEs. Our analysis of this unified transposition activity dataset reveals that the Tc1/*mariner* superfamily elements, particularly those showing high copy number and possessing an IR-DR structure, are active. The high activity contributes to

pervasive horizontal transfers of Tc1/*mariner*.^{44–46} Furthermore, compared with other TEs, active Tc1 elements are less associated with MITEs. In addition to activity and evolutionary dynamics, active TEs also exhibit diversified application features, such as the cargo capacity and insertion patterns. We finally demonstrated that the TE with the highest activity, i.e., *Mari-ner2_AG* (MAG), largely outperformed the conventional lentiviral vector and SB100X used in CAR-T therapy against both hematological malignancies and solid tumors. Overall, our study not only highlights the divergence of DNA TEs in terms of transposition activity and evolutionary dynamics but also expands TE-based genome engineering toolbox by incorporating functionally diverse TEs, particularly the highly potent MAG.

RESULTS

One hundred and thirty active DNA TEs were predicted across 102 metazoan genomes

To increase the success rate of identifying DNA TEs mobile in human cells, we began with a literature survey and found 20 TEs active in mammalian cells that were cumulatively identified in past decades (Table S1; STAR Methods). With the exception of Ac from plants, the other 19 TEs were from metazoans, with fish at the top contributing six TEs (e.g., SB from salmon³¹). Thus, we searched for active TEs in 100 publicly available metazoan genomes with TE annotation, along with two fish genomes without TE annotation. For each active TE candidate, we annotated open reading frame (ORF) and TIR. In addition, since DNA TEs recognize specific genomic target sites,^{25,47} we also extracted target information. These three types of information would enable the downstream experimental survey.

Specifically, from all 100 annotated genomes hosted by the UCSC Genome Browser,⁴⁸ we predicted 124 potentially active DNA TEs via two rounds of filtration (Figure 1A; STAR Methods). First, from 2,024 distinct DNA TEs corresponding to millions of copies harbored by these genomes, we extracted 1,577 TEs with at least one potentially full-length autonomous copy defined according to the consensus sequences annotated by Repbase.⁴⁹ We then only retained candidates harboring TIRs and ORFs with a transposase domain in the consensus sequences. After addressing redundancy caused by TEs shared across multiple species (STAR Methods), we identified 124 unique candidates (Table S2A). We further annotated target sites by searching the hallmark sequence of transposition, i.e., target site duplication (TSD), which comprises a short duplicated sequence flanking TIRs.¹²

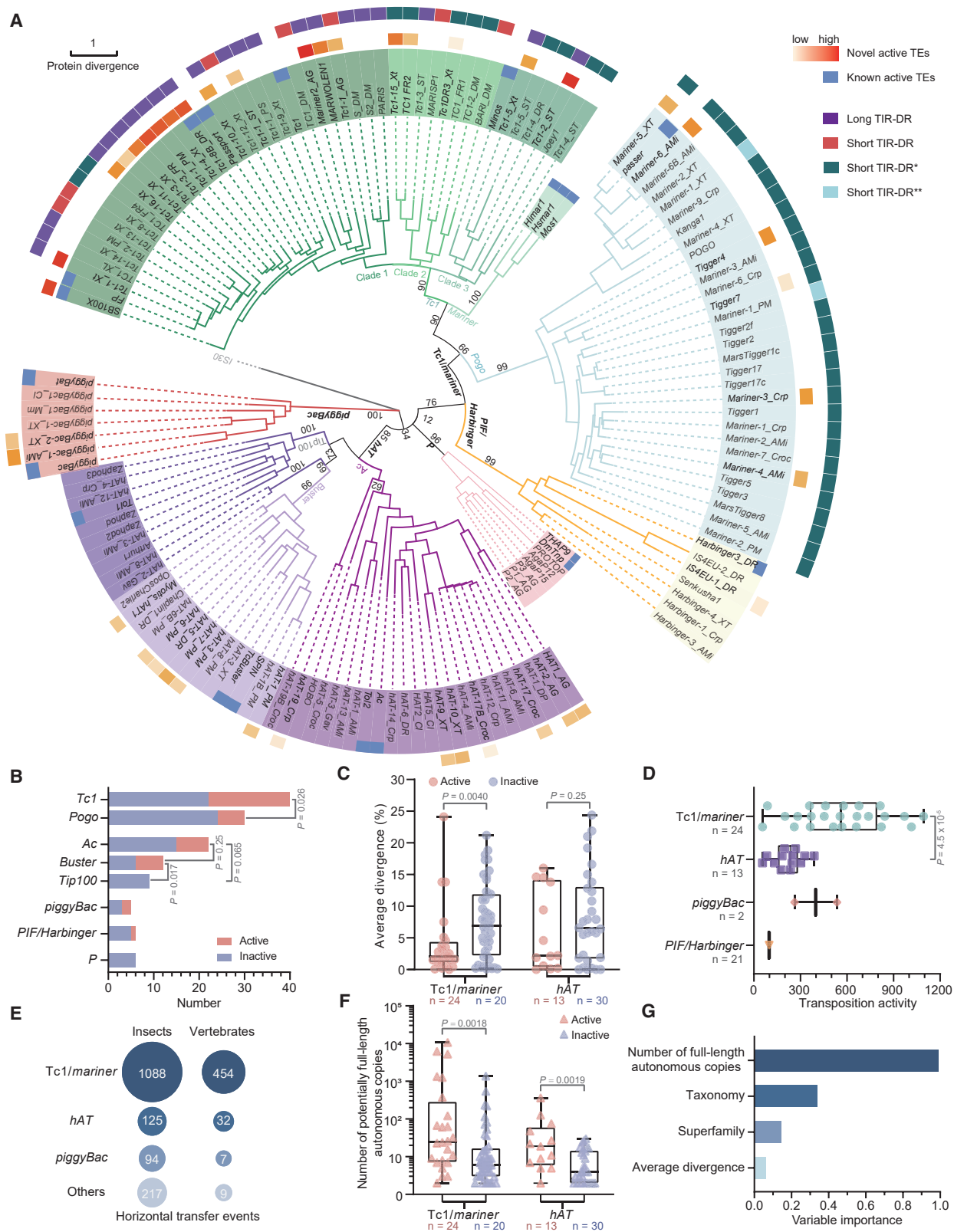
To expand our survey to genomes without TE annotation, we searched for active DNA TEs in two additional fish genomes,

(C) The distribution of the average percent divergence of DNA TEs relative to the consensus. Ascidiacea and Echinoidea in (B) are excluded due to a small number of TEs. The boxplot shows the median with a black line, the first and third quartiles with hinges, and the minimum and maximum with whiskers. The actual data points are also overlaid.

(D) The scheme for quantifying the transposition activity of TEs. For *PIF/Harbinger* elements, a second helper plasmid encoding the myb-like protein was co-transfected. For plasmids, PGK (Phosphoglycerate Kinase) and CMV (Cytomegalovirus) represent two distinct promoters, while *puroR* indicates a gene with puromycin resistance gene and GFP indicates the green fluorescent protein.

(E) The transposition activity of 40 active TEs and SB100X in HEK293T cells. For each superfamily, the number of active TEs and total candidates is shown in parentheses. Two TEs from the unannotated fish genomes are shown in purple. Data are represented as mean $\pm$ SD across two biological replicates.

See also Figure S1.



(legend on next page)

Sinocyclocheilus tileiornes (cave fish) and *Pseudoliparis swirei* (snailfish in the deep sea). In addition to the aforementioned fact that fish contribute more active DNA TEs (Table S1), the genomes of these two species from stressful environments may harbor an increased number of TEs.⁵⁰ After *de novo* TE annotation, including TE consensus construction with RepeatModeler and TE search with RepeatMasker (STAR Methods), we performed the same two rounds of filtration (Figure 1A). By correcting low-quality consensus sequences generated by RepeatModeler and precisely predicting TIRs, we identified six final candidates (Table S2B). Thus, we compiled a merged candidate list of 130 DNA TEs.

The distribution of TEs with respect to species and superfamilies was consistent with previous knowledge, indicating a high quality of the final candidate dataset. Although mammals accounted for 55.9% (57/102) of the included species (Figure S1A), TE candidates from mammals constituted only 15.4% (20/130, Figure 1B) of all candidates and tended to be old, as indicated by high percentage divergence relative to the consensus sequence approximating the ancestral sequence (median: 14.6%, Figure 1C). These patterns are consistent with the loss of DNA TE activity over at least the last 40 million years in most mammal lineages.^{51,52} The top two species, *Alligator mississippiensis* and *Xenopus tropicalis*, harbored 55 (42.3%) candidates, in line with the high abundance of DNA TEs in these two species.^{53–55} On the other hand, 130 candidates span five superfamilies (Tc1/*mariner*, *hAT*, *piggyBac*, *PIF/Harbinger*, and *P*), with Tc1/*mariner* and *hAT* being the top two in terms of abundance, consisting of 70 and 43 candidates, respectively. This pattern is again in line with a previous report showing that the Tc1/*mariner* and *hAT* elements are most pervasive in metazoans.⁵⁴ Among the six *P* elements, five were from the *Anopheles gambiae* genome, and one was from the *Drosophila melanogaster* genome, which was also consistent with the insect bias of this superfamily.^{56,57} Notably, all 20 known cases were from the same five superfamilies (Table S1).

Altogether, by developing two bioinformatic pipelines, we predicted 130 active TE candidates and annotated their coding sequences, TIRs, and TSDs.

Functional screening in human cells identified 40 active TEs

To evaluate the transposition activity of the 130 TEs in human cells, we implemented a well-established assay,^{31,58,59} co-transfecting one helper plasmid and one donor plasmid (Tables S3A and S3B). Specifically, the helper encodes a human codon-opti-

mized transposase (Figure 1D). The donor encodes puromycin resistance gene and a GFP reporter flanked by corresponding noncoding sequences (TSDs, TIRs, and spacer sequences between TIRs and coding regions, Figure 1A). After transfection followed by puromycin selection, the number of resistant colonies represents the transposition activity. We profiled 60 TEs in HEK293T and HeLa cell lines (STAR Methods). Active TEs were largely shared by both cell lines, and the activity level was correlated (Spearman's $\rho = 0.83$, Figures S1B and S1C). We therefore screened the remaining 70 candidates in only HEK293T cells.

Co-transfection assays identified 40 (30.7%) active TEs (Figure 1E; STAR Methods). Among these TEs, three showed higher activity than the positive control (SB100X), and the most active *Mariner2-AG* show 19.8% higher activity compared with SB100X with the difference being statistically significant (Student's *t* test $p = 0.016$). The second most active TE (*Tc1-2-ST*) was from the cave fish and obtained from *de novo* annotation.

Collectively, by quantifying TE transposition activity in human cells, we identified 40 active TEs, including the highly active *Mariner2-AG*.

Forty active TEs increase the evolutionary diversity of the DNA TE toolbox

Given the 40 active DNA TEs and 90 inactive ones, we wondered how these TEs increased phylogenetic breadth relative to the 20 known active TEs (Table S1).

We first anchored all TEs in a phylogenetic tree, reconstructed on the basis of alignment of the conserved catalytic DDE/D domain (STAR Methods). Consistent with a previous cross-superfamily analysis,¹² the relationships between superfamilies were not well resolved, as indicated by a low confidence score (bootstrap < 70, Figure 2A). However, the within-superfamily phylogeny was resolved, in which all members were clustered with members from the same superfamily. Moreover, the established phylogeny of three families in the animal Tc1/*mariner* superfamily, namely, *Tc1*, *Mariner*, and *Pogo*, was reproduced.¹¹ Similarly, the phylogeny of three families in the *hAT* superfamily, *Ac*, *Buster*, and *Tip100*, was also largely reproduced,^{15,60} although two annotated *Tip100* elements were clustered into the *Buster* branch. This ambiguity possibly reflects the controversy about whether *Tip100* represents a family separated from *Ac* and *Buster*.^{61–63}

These 40 active DNA TEs not only covered most major branches on which known active TEs were located but also expanded the phylogenetic space, especially for the Tc1/*mariner*

Figure 2. Evolutionary analyses of active and inactive DNA TEs

(A) Phylogenetic tree of the 130 candidate TEs and the 20 known active TEs. TEs in black and in gray are active and inactive, respectively. The bootstrap confidence scores are labeled along major branches representing families or superfamilies. From inside to outside, the three concentric squares indicate known TEs, novel active TEs, and TIR types (see also Figure 3A). Notably, although we focused on novel TEs (STAR Methods), *Tc1-8B_DR* was discovered in the process of our project.⁵⁹

(B) The distribution of active TEs across superfamilies or families.

(C) The distribution of average divergences across active and inactive TEs. For (C), (D), and (F), the convention follows Figure 1C.

(D) The distribution of the transposition activity across superfamilies.

(E) Horizontal transfers of DNA TEs in insects and vertebrates, broken down per superfamily. The data are from two studies.^{45,64}

(F) The distribution of potentially full-length autonomous TE copy number.

(G) The weight of evolutionary features in predicting the presence or absence of transposition activity in human cells.

See also Figure S1.

and *hAT* superfamilies (Figures 1B and 2A). Within the *Tc1* family of the *Tc1/mariner* superfamily, our dataset contributes 17 cases, in addition to the six previously known cases. These 23 cases are distributed in three clades, with clade 1 showing the highest abundance, including aforementioned *Mariner2-AG* and *SB100X* (Figure 2A). *Tc1-15_Xt*, *TC1_FR2*, and *Tc1DR3_Xt* are scattered in clade 2, which does not include any previously known active TE. The remaining clade 3 includes two active TEs, i.e., *Tc1-2_ST* and *Tc1-5_Xt*. Compared with the *Tc1* family, the *Pogo* family has not been experimentally explored until the recent identification of *Passer*.³⁴ We have now added six cases to this family, representing a broad phylogenetic space (Figure 2A). Our screen did not include any candidate from the *Mariner* family, consisting of three known cases. This depletion is possibly because of the distribution bias of *Mariner* in arthropods,^{21,65} while mammals were overrepresented in our species list (Figure S1A). Within the *hAT* superfamily, only five known cases are scattered across all three families, *Ac*, *Buster*, and *Tip100*, while 13 newly discovered cases with moderate sequence similarity (<42%) to known cases are distributed in the *Ac* and *Buster* families.

For the remaining three superfamilies with fewer cases tested (Figure 1B), the increase in diversity was less pronounced. For the *piggyBac* and *PIF/Harbinger* superfamilies, only two known cases and one known case existed, respectively (Figure 2A). Two active *piggyBac* elements and one *PIF/Harbinger* element are now added, all of which show low similarity relative to the known counterparts (<15%). In addition, although two known active *P* elements have been reported (Figure 2A),⁵⁶ TEs of this superfamily are believed to be mainly active in insects since their transposition likely depends on insect-specific partner proteins.^{31,66,67} Thus, the absence of *P* activity in human cells across all six candidates possibly reflects a general pattern.

In summary, our screening substantially increases the number and diversity of active DNA TEs.

Tc1 elements with high copy numbers are more likely to be active than other TEs

Motivated by the apparent enrichment of active DNA TEs in the *Tc1* family (Figure 2A), we analyzed the distribution of active TEs across superfamilies and families. The overall proportion of active TEs was 34.2% for *Tc1/mariner*, 30.2% for *hAT*, 40% for *piggyBac*, 16.7% for *PIF/Harbinger*, and 0 for *P* superfamily (Figure 1E). Since the total number of candidate TEs in the last three superfamilies was small (5, 6, and 6), these proportions were associated with uncertainty. By exploiting the large number of *Tc1/mariner* and *hAT* candidates, we performed family-level analyses (Figure 2B). We found that a significantly higher proportion of *Tc1* elements than *Pogo* elements were active (45% vs. 20%, Fisher's exact test $p = 0.026$, Figure 2B), and *Buster* (50%) elements were the most active, followed by *Ac* (31.8%) and *Tip100* elements (0%), with the difference between *Buster* and *Tip100* being significant ($p = 0.017$). These differences could be attributed to the fact that DNA TEs from more active families tend to be evolutionarily younger, as approximated by divergence relative to the consensus sequence (Mann-Whitney test $p < 0.05$, Figures S1D and S1E). Consistently, both active *Tc1/mariner* elements and *hAT* elements tend to show lower diver-

gence than their inactive counterparts (median: 2.1% vs. 6.9%, 2.2% vs. 6.5%, Figure 2C), although only the former comparison is statistically significant due to the larger sample size (Mann-Whitney test $p = 0.004$).

For active DNA TEs, we examined the distribution of transposition activity. *Tc1/mariner* showed the highest median activity (561), followed by *piggyBac* (399), *hAT* (236), and *PIF/Harbinger* (100, Figure 2D). The comparison between the two superfamilies with a relatively higher number of active TEs, namely *Tc1/mariner* and *hAT*, showed statistical significance. For family-level comparison, the median activity of *Tc1* was marginally significantly higher than that of *Pogo* (662 vs. 441, Mann-Whitney test $p = 0.088$, Figure S1F), while the median activity of *Ac* and *Buster* was similar (249 vs. 208). The high heterologous transposition activity of *Tc1/mariner* elements could contribute to their horizontal transfers.^{44–46,64} Consistently, the *Tc1/mariner* superfamily accounted for more than 70% of horizontal transfers in both insects and vertebrates (Figure 2E; STAR Methods). In addition to factors such as superfamily/family and divergence (Figures 2B–2D and S1D–S1F), active DNA TEs are also associated with a high number of potentially full-length autonomous copies. The difference is significant for both *Tc1/mariner* and *hAT* superfamilies (Mann-Whitney test $p = 0.0018$ and 0.0019, respectively, Figure 2F).

Clearly, all these evolutionary features could overlap with each other; for example, TEs with high copy numbers also show low divergence. Moreover, homologous TEs should not be viewed as independent data points. To control these technical issues and generate an overview on which evolutionary information best predicts whether DNA TEs are active in human cells, we performed a joint modeling while controlling homology between TEs (STAR Methods). We found that copy number was the strongest predictive factor, while the other factors only made moderate contributions (Figure 2G). This result is unsurprising since copy number most directly reflects recent transposition activity in the native genomes of TEs.

Collectively, our between- and within-superfamily analyses show that high-copy-number *Tc1* elements are most active in human cells.

Tc1 elements exhibit higher activity in the presence of an IR-DR structure and show a lower propensity for evolving into MITEs

Next, by exploiting the unified activity dataset (Figure 2A), performing dedicated experiments, and analyzing MITE sequences, we estimated the importance of noncoding sequences for transposition activities. We focused on the *Tc1/mariner* superfamily since it has the largest number of tested cases (Figure 2A) and IR-DR structure has been characterized in this superfamily.^{11,17–19}

We first found that active *Tc1/mariner* elements were more often associated with TIRs harboring IR-DR structure. Specifically, in line with the diversity of TIRs of *Tc1* elements,^{68,69} sequence analyses revealed four distinct IR-DR architectures (Figure 3A; Table S4; STAR Methods): (1) long TIR-DR: TIRs with a length exceeding 100 bp, characterized by the presence of two pairs of DRs similar to SB^{11,17,18}; (2) short TIR-DR: TIRs with a length shorter than 100 bp, harboring one pair of DRs

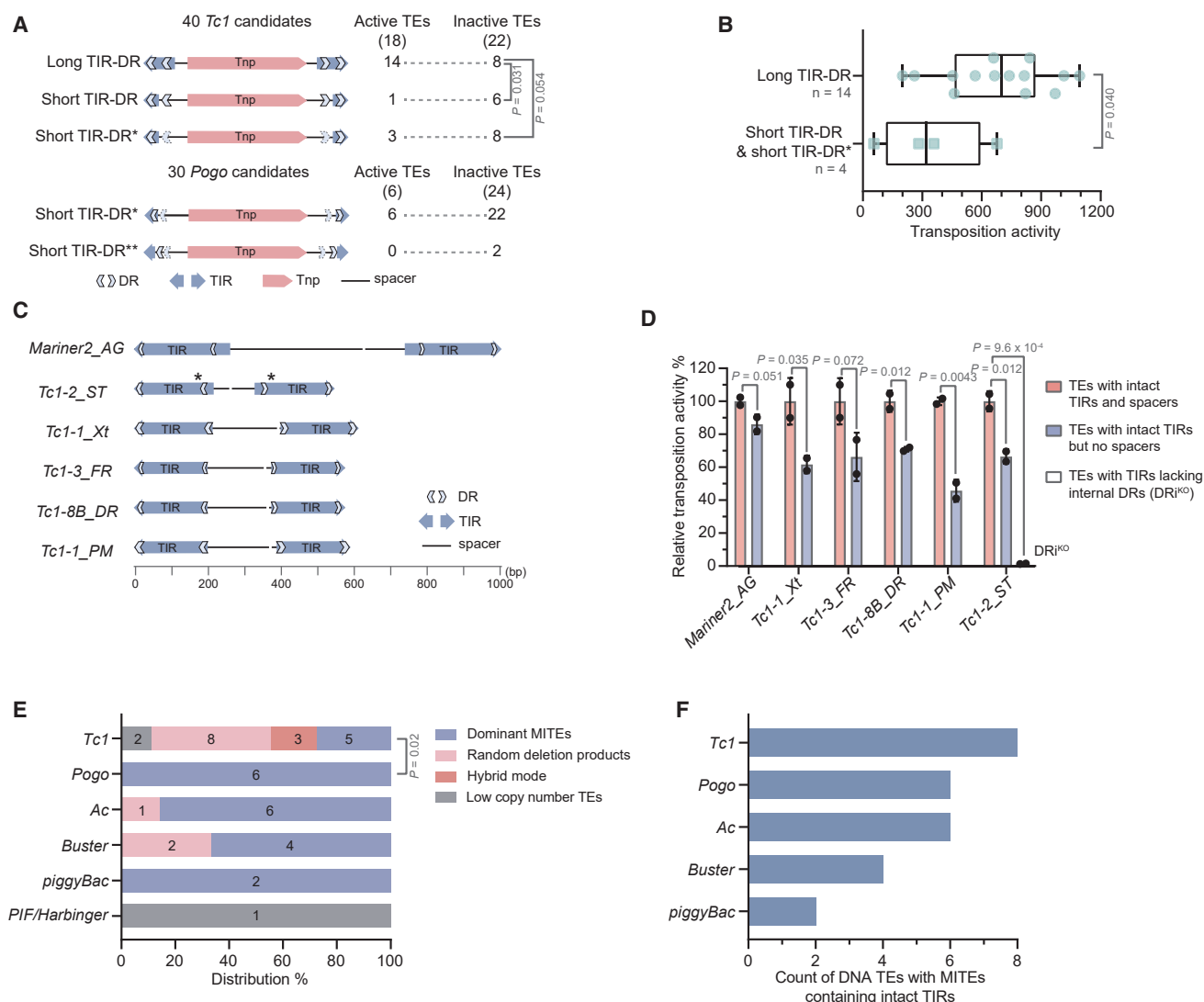


Figure 3. Evaluation of the importance of noncoding sequences

(A) Four TIR architectures in terms of DRs.

(B) The distribution of transposition activity across active *Tc1* elements. The convention follows Figure 1C.

(C) A schematic illustration of the transposon fragment in the top six active *Tc1* elements. The breaking point of spacers indicates the previous position of transposase.

(D) The relative transposition activity of the top six active *Tc1* elements with spacer or internal DRs removed. DRi^{KO} represents the deletion of two inner DRs in *Tc1-2_ST* (the black asterisks in C). Data are represented as mean $\pm$ SD across two biological replicates.

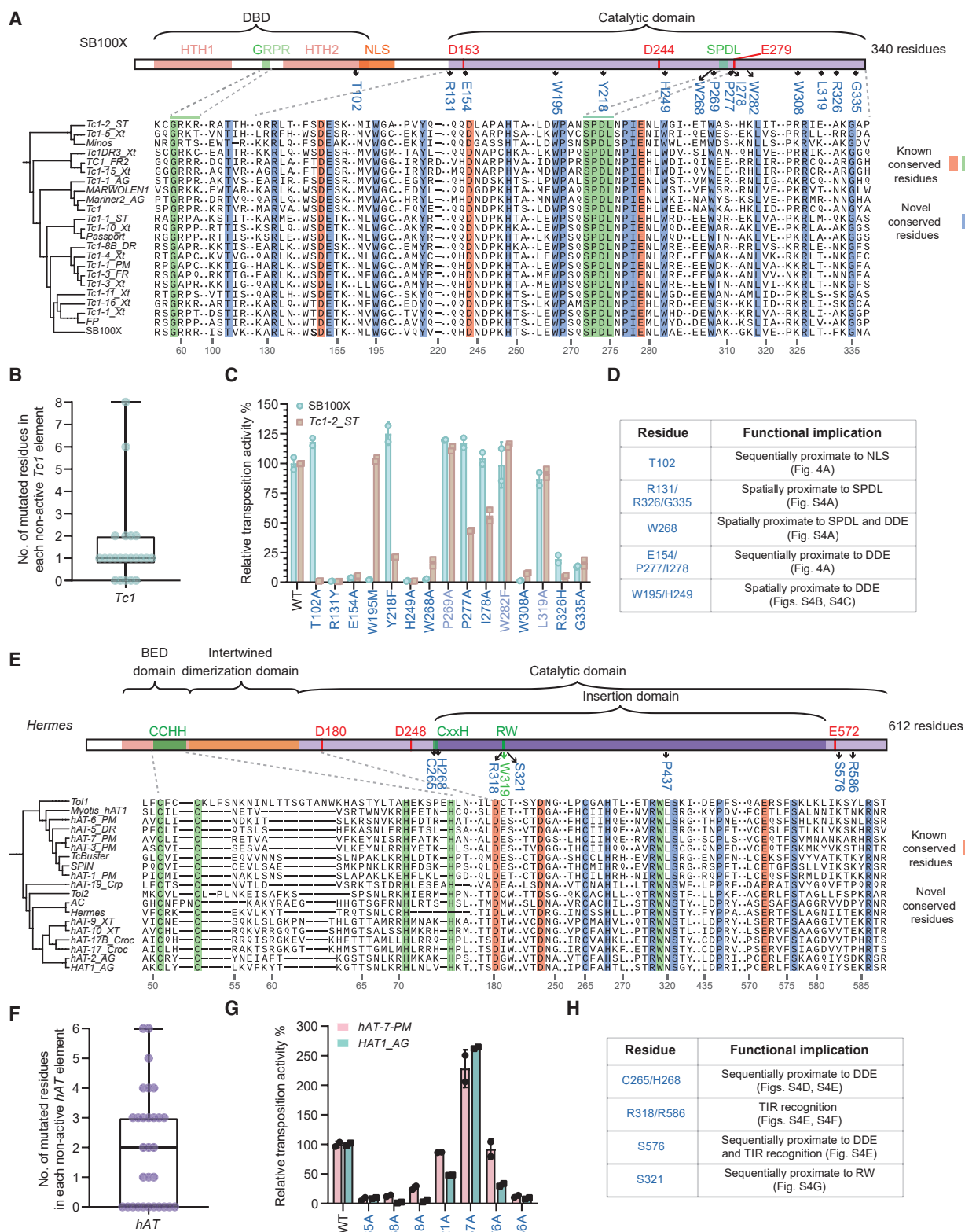
(E) The distribution of 40 active TEs in terms of evolutionary models, broken down into superfamilies or families. Low copy number TEs: ≤ 3 non-autonomous copies.

(F) Count of DNA TEs with MITEs containing intact TIRs.

See also Figure S2.

within the TIRs and the other pair in the spacer region; (3) short TIR-DR*: similar to the short TIR-DR architecture, but with one DR missing from the spacer region; and (4) short TIR-DR**: similar to the short TIR-DR* architecture, but with all three DRs present in the spacer region. *Tc1* elements match the first three architectures, while *Pogo* elements only match the last two architectures (Figure 3A). This is in line with the fact that *Pogo* TIRs are generally short (<40 bp)^{34,70}. Within *Tc1* family, members with standard IR-DR structures or long TIR-DR architec-

tures are more frequently active (14/22 or 63.7%, Figure 3A) than members with short TIRs (14.3%–27.2%, Fisher's exact test $p \leq 0.054$). Furthermore, among active *Tc1* elements, those with long TIRs showed higher activity (median: 700 vs. 300, Mann-Whitney test $p = 0.040$, Figure 3B). Notably, after accounting for phylogenetic dependence or homology, the former comparison becomes more pronounced (phylogenetic test $p \leq 0.033$; STAR Methods), whereas the latter shows no significant difference ($p = 0.293$).



(legend on next page)

In addition to TIRs or IR-DRs, the spacers of DNA TEs may also affect transposition activity.^{28,71–73} To test this hypothesis, we compared the vector with a spacer and the counterpart lacking a spacer (Figures 1A and 3C) for the top six most active TEs, all of which were *Tc1* elements (including *Mariner2_AG*, *Tc1-2_ST*, *Tc1-1_Xt*, *Tc1-3_FR*, *Tc1-8B_DR*, and *Tc1-1_PM*; Figure 1E; Table S3). Despite the phylogenetic divergence among these TEs (Figure 2A), all of them showed a moderate decrease in transposition activity, with four reaching statistical significance (Figure 3D), indicating that *Tc1* spacers generally harbor functional elements underlying the activity. For *Tc1-2_ST*, which was phylogenetically diverged from the other five elements (Figure 2A), we further removed the two internal DRs and observed the complete loss of activity (Figures 3C and 3D). Thus, for IR-DR *Tc1* elements belonging to either clade 1 (e.g., SB100X^{18,74} in Figure 2A) or clade 3 (*Tc1-2_ST*), DRs are often essential.

Since TIRs are essential for transposition,^{14,20,29,75} we expected that MITEs, if present, would retain the intact TIRs of their corresponding autonomous TEs. Hence, for 40 active DNA TEs, we analyzed their MITEs (non-autonomous DNA TEs with high copy numbers; STAR Methods). We identified three scenarios (Figure S2A; Table S4): first, as described by Pace et al.,²⁶ one active autonomous DNA TE corresponds to a few MITEs; second, no MITE is present where non-autonomous TEs represent random low copy number deletion derivatives of autonomous TEs; and third, in a hybrid mode, MITEs contribute a proportion of non-autonomous TEs, and the remaining non-autonomous TEs are deletion derivatives with low copy numbers. Among different families or superfamilies, most active TEs largely fit the first scenario, except for *Tc1* elements, among which 8 (44.4%) and 3 (18.8%) fit the second and third scenarios, respectively (Figure 3E). Regardless of the first or third scenario, all MITEs always harbor intact TIRs of active DNA TEs (Figure 3F), supporting their ubiquitous importance.

In summary, our findings corroborate the importance of TIRs. Moreover, regarding *Tc1* elements, they are more likely to be active when long TIRs or IR-DR structures are present, their spacers affect activity, and they are less likely to be associated with MITEs.

Alignments of active DNA TEs along with experimental validations reveal essential amino acids

In parallel to analyses of noncoding sequences, comparisons between active and inactive DNA TEs also reveal key protein residues underlying transposition activity.

We first focused on the *Tc1* family within the *Tc1/mariner* superfamily due to the abundance of both active and inactive elements in this group (Figure 2A). We formulated three interconnected hypotheses: (1) conserved amino acids, such as D, D, and E/D,^{11,76} are shared among active DNA TEs; (2) inactive TEs exhibit greater variability at these sites; and (3) substitutions of these amino acids affect transposition activity. Consistently, sequence alignment revealed 23 conserved sites shared by all 23 active TEs, including eight previously experimentally validated essential sites (e.g., DDE), along with 15 novel sites (Figure 4A; STAR Methods). Conversely, the inactive TEs showed a median of one mutation at these sites (Figure 4B). To ascertain the functional impact of these conserved sites, particularly the 15 novel sites, we introduced mutations resulting in a change to either the amino acid state found in inactive TEs or alanine (Ala) in cases where multiple derived states were present in inactive TEs (Table S3C; STAR Methods). To gain a comprehensive understanding, we examined two phylogenetically representative members, SB100X and *Tc1-2_ST* (Figure 2A). At seven sites (e.g., R131), introduced mutations largely abolished the activity of both SB100X and *Tc1-2_ST* (Figure 4C). Mutations at five sites (e.g., T102) disrupted the activity of either SB100X or *Tc1-2_ST*, indicating differential tolerance. Three sites (e.g., P269) tolerated mutations in both *Tc1* elements, suggesting that these sites affect other functions (e.g., insertion preference) rather than transposition activity. For 12 sites involved in activity, structural modeling revealed that 10 sites were sequentially or spatially close to known critical motifs, including DDE, SPDL (S for serine, P for proline, D for aspartic acid, and L for leucine),⁷⁷ and nuclear localization signal (Figures 4D and S2B–S2D). The functional roles of the remaining two sites are more elusive.

Although we could not perform a similar analysis for the *Pogo* family due to the small number of active TEs, we analyzed the *hAT* superfamily. All patterns were reproduced: (1) active TEs possessed the CCHH motif (C for cysteine and H for histidine), DDE motif, and experimentally validated W319 of the RW motif (R for arginine and W for tryptophan),¹⁵ along with seven novel sites (Figure 4E); (2) inactive TEs harbored a median of two mutations across these 15 sites (Figure 4F); (3) experimental investigation revealed the essentiality of four sites for two diverged *hAT* members (*hAT-7_PM* and *HAT1_AG*, Figure 2A), together with the differential importance of two sites (Figure 4G); and (4) these six sites were situated in the spatial proximity to DDE, the TIR interaction region, or the sequential vicinity of

Figure 4. Evaluation of the functional importance of conserved residues

- (A) Conserved (unchanged) residues of active *Tc1* elements. Domains or motifs of SB100X are shown at the top, while the position is numbered on the basis of SB100X at the bottom. The tree topology is based on Figure 2A.
- (B) The distribution of substituted residue number across inactive *Tc1* elements. This panel and (F) follow the figure convention of Figure 1C.
- (C) Mutational effects on transposition activity of *Tc1* elements in HEK293T cells. Data are represented as mean $\pm$ SD across two biological replicates. The three residues tolerating mutations for both TEs are marked in light blue.
- (D) Proximity of novel functional residues relative to known essential motifs of SB100X transposase.
- (E) Conserved (unchanged) residues of active *hAT* elements. Domains or motifs of *Hermes* transposase are shown at the top, while the position is numbered on the basis of *Hermes* transposase at the bottom. The tree topology is based on Figure 2A.
- (F) The distribution of substituted residue number across inactive *hAT* elements.
- (G) Mutational effects on transposition activity of *hAT* elements in HEK293T cells. Data are represented as mean $\pm$ SD across two biological replicates.
- (H) Proximity of novel functional residues relative to known essential motifs of *Hermes* transposase.

See also Figures S2 and S7.

the RW motif (Figures 4H and S2E–S2H). Unexpectedly, the remaining P437A mutation increased the transposition activity, suggesting that P437 is evolutionarily maintained to fine-tune the activity.

Altogether, the *in silico* and experimental analyses of the *Tc1* family and the *hAT* superfamily revealed that our activity data not only recovered those known essential amino acids but also revealed novel important sites.

Active DNA TEs show diverse functional features

To evaluate the application potential of these active TEs, we further studied three critical features: the insertion preference, cargo size tolerance, and overproduction inhibition (transposition inhibition upon high expression^{34,43,78}). We focused on the most active and phylogenetically representative DNA TEs, including the top six active *Tc1* elements, one *Pogo* elements (*Tigger4*), *hAT-7_PM*, *HAT1_AG*, and *piggyBac-1_AMi* (Figure 2A).

We analyzed the insertion patterns of these TEs together with SB100X and *piggyBac* as controls (Table S5; STAR Methods). Given rich functional annotations of K562 cells,⁷⁹ we performed analyses in this cell line after ensuring the activity of all 12 TEs (Figure S3A). DNA TEs have evolved three insertion profiles: insertion into transcriptionally active regions to ensure their own expression; insertion into less important regions, or so-called safe harbor, to make them less harmful to the host; and semi-random insertion.^{80–82} Mapping data across 12 TEs demonstrated all three profiles (Figures 5A and S3B–S3D): (1) *Tigger4*, *hAT* members, and *piggyBac* showed enrichment toward transcriptional start sites, with *HAT1_AG* showing the strongest enrichment and *Tigger4* showing the weakest enrichment; (2) *piggyBac-1_AMi* insertion was biased toward both repressed chromatin and various safe harbors, such as regions without cancer-related genes; and (3) the remaining seven *Tc1* members, including SB100X, exhibited a semi-random distribution with a weak enrichment toward genic regions. These patterns align with the established bias of *piggyBac* toward transcriptional start sites and the semi-random pattern of SB100X.^{83–86} Additionally, an analysis across histone marks also produced largely consistent patterns, such as the enrichment of *Tigger4*, *hAT* members, and *piggyBac* toward active marks (Figure S3E).

Furthermore, with the exception of *Mariner2_AG*, which possesses a small number of native sites, the remaining nine TEs recognize similar motifs in human cells and their respective native genomes (Figure S3F): (1) five *Tc1* elements together with *Tigger4* (a *Pogo* element) recognized a TA motif; (2) *HA-T1_AG*, an *Ac* element, bound to an 8-bp motif in which T and A were overrepresented at the second and seventh positions, respectively, while *hAT-7_PM*, a *Buster* element, bound to a different 8-bp motif in which T and A were overrepresented at the fourth and fifth positions, respectively; and (3) *piggyBac-1_AMi* recognized a TTAA motif. All of these motifs were consistent with the previous reports spanning different families or superfamilies.^{12,15,43,85,87}

Second, with SB100X as a control, we evaluated the cargo size tolerance capability for 9 out of the 10 aforementioned DNA TEs, excluding *Tigger4* due to its relatively low activity (Figure 1E; STAR Methods). All TEs display their peak activity with the smallest

cargo size (2 kb, Figure 5B), aligning with previous reports that smaller non-autonomous DNA TEs tend to show higher activity compared with their autonomous counterparts.^{28,71} However, *Mariner2_AG*, *Tc1-2_ST*, and *Tc1-1_Xt* showed relatively stronger tolerance for larger cargos, with their transposition activity remaining at approximately 30% even with a 10-kb cargo, which is comparable to the extensively optimized SB100X.

Finally, we assayed the overproduction inhibition of the six active *Tc1* elements together with SB100X as the control. To be less deleterious to hosts, some DNA TEs self-regulate their transposition; for instance, high levels of transposase lead to protein aggregation and subsequent reduction in transposition activity.^{88,89} By contrast, under an increasing transposase level, the activity of other TEs can either reach a plateau or monotonically rise.^{73,90} Six *Tc1* elements together with SB100X seemed to show all three possibilities. First, consistent with a previous report,⁵⁹ SB100X showed overproduction inhibition, where transposition activity reached a peak with the increasing doses of the helper plasmid (encoding transposase) and declined with the further increases of dose (Figure 5C). *Tc1-3_FR*, *Tc1-1_PM*, and *Tc1-1_Xt* seemed to be also subjected to overproduction inhibition with varying dynamics. By contrast, *Tc1-2_ST* reached a plateau, while the remaining two *Tc1* members showed a roughly linear increase in activity along with an increase of transposase. Certainly, these three may also show overproduction inhibition with further increases of helper doses.

Collectively, our head-to-head assays highlight the functional diversity of the active DNA TEs, indicating broad application potentials.

CAR-T cells engineered with *Mariner2_AG* show high efficacy in eradicating tumors

Finally, to explore the practical utility of these active DNA TEs, we evaluated them in therapeutic applications, specifically TE-based CAR-T cell therapy.

Similar to our approach in HEK293T cells (Figure 1D; STAR Methods), we first identified the most efficient TE vector in human T cells. Specifically, we tested *Mariner2_AG*, *Tc1-1_Xt*, *Tc1-3_FR*, and *Tc1-8B_DR* on the basis of their activity, insertion randomness, and efficiency across different cargo sizes (Figures 1E, 5A, and 5B). Among them, *Mariner2_AG* (subsequently referred to as MAG) exhibited the highest activity (16% vs. 2%–12%, Mann-Whitney test $p < 0.01$, Figure 6A). For comparison, SB100X and a lentiviral vector were included for their routine usage in CAR-T therapies.^{39–41} Notably, among the two widely used lentiviral vectors, FUW^{91,92} and pRRLSIN,^{93–95} we selected FUW due to its better performance in tumor cell lysis (Figures S4A–S4C).

A typical CAR-T experiment consists of CAR-T cell generation and amplification together with functional evaluation *in vitro* and *in vivo* (Figure 6B). With electroporation, we co-delivered two plasmids harboring the CD19 CAR and MAG transposase into T cells. Notably, although the electroporation causes cell toxicity,³⁹ the cell numbers of all samples were similar to that of the untreated control after recovery (Figure S4D; STAR Methods). Despite the different vectors used, the proportions of CAR-positive T cells generated with MAG (MAG-CD19),

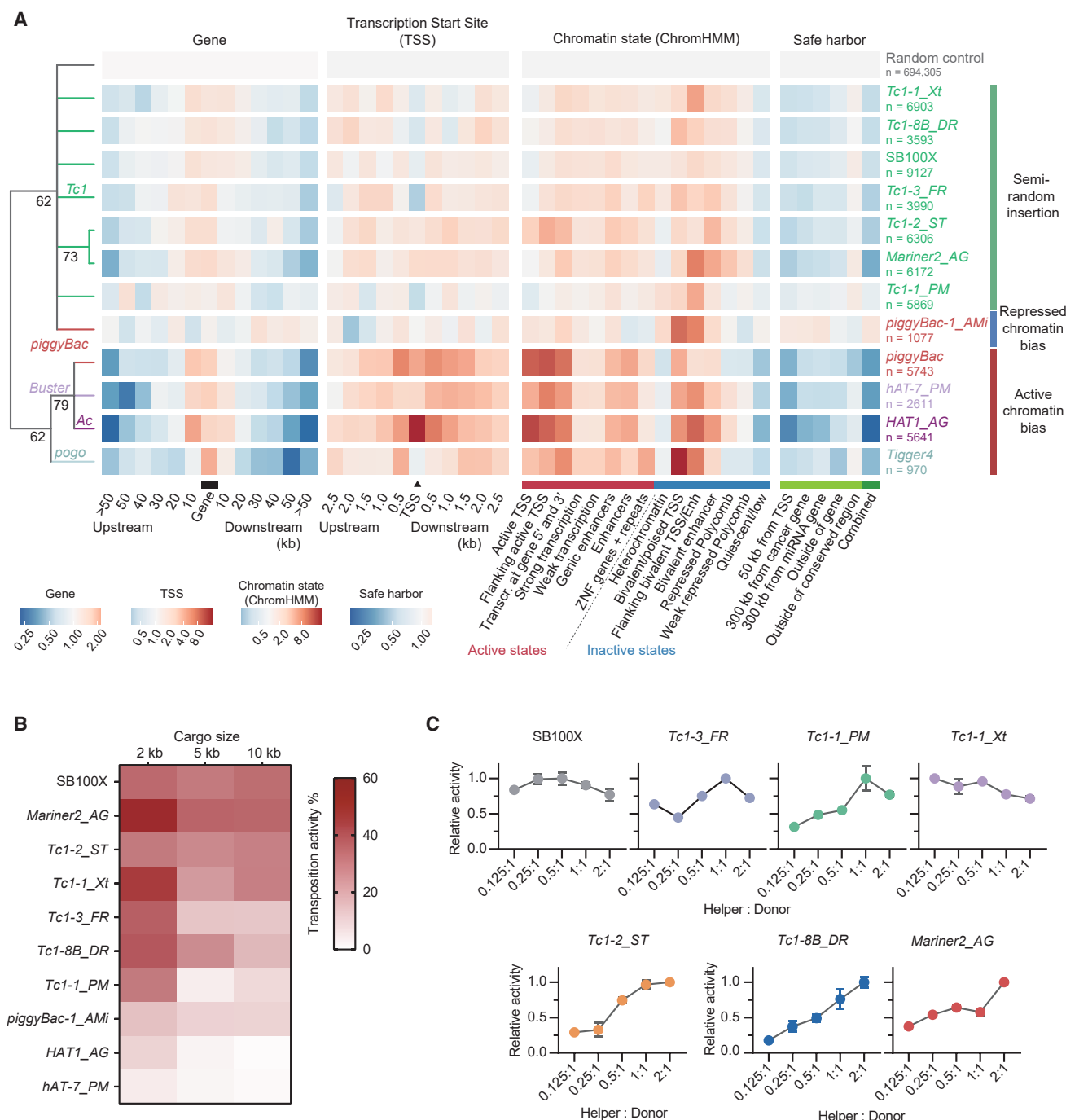


Figure 5. Functional characterization of active TEs

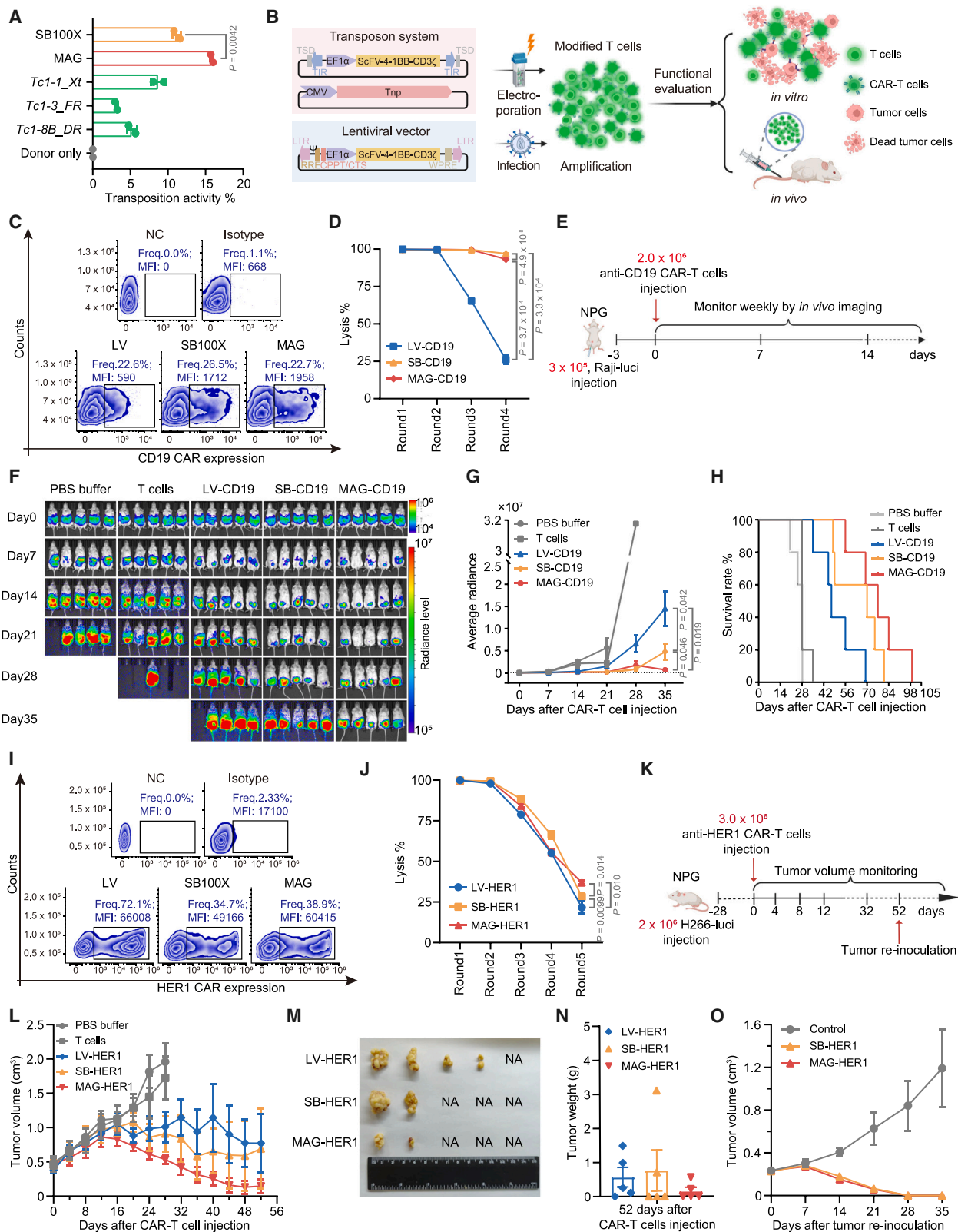
(A) The integration profile heatmap in K562 cells. The dendrogram on the left was generated on the basis of the Euclidean distances between rows after log transformation. Branches with bootstrap confidence scores lower than 60 are collapsed.

(B) The transposition activity of TEs with different cargo sizes in HEK293T cells.

(C) The relative transposition activity of TEs at various helper-to-donor ratios in HeLa cells. Data are represented as mean $\pm$ SD across two biological replicates. See also Figures S3 and S7.

SB100X (SB-CD19), or lentivirus (LV-CD19) were similar (22%–27%, Figure 6C). Next, we assessed the *in vitro* functionality of CAR-T cells, using a hematological cancer cell line (K562) over-

expressing CD19 (STAR Methods). All three types of CAR-T cells showed similar abilities to eliminate tumor cells across different effector-to-target (CAR-T cells vs. cancer cells) ratios



(legend on next page)

(Figure S4E). However, after subjecting the cells to four rounds of tumor cell challenges, which induce T cell exhaustion,⁹² the LV-CD19 CAR-T cells exhibited a decline in cytotoxicity, showing only approximately 25% tumor lysis in the last round (Figure 6D). By contrast, MAG-CD19 and SB-CD19 CAR-T cells maintained a high efficacy, erasing almost all tumor cells even in the last round. Given the similar T cell subtype composition observed throughout this process (Figure S4F), the enhanced efficacy of MAG-CD19 and SB-CD19 cells may stem from a greater proportion of CAR-positive cells following each challenge round compared with LV-CD19 cells (Figure S4G).

To evaluate the efficacy *in vivo*, we established a tumor cell line-derived xenograft (CDX) model by injecting Raji-luciferase cells (a hematological cancer cell line expressing luciferase) into mice, followed by the administration of CAR-T cells (Figure 6E). LV-CD19 reduced tumor compared with the phosphate-buffered saline (PBS) buffer and T cell control groups. However, SB-CD19 exhibited even better performance, and MAG-CD19 outperformed both with respect to control of tumor volumes (Figures 6F and 6G) and overall survivals (Figure 6H). To validate MAG-CD19's superiority over LV-CD19 and SB-CD19, we conducted tests under the more challenging condition, using an increased number of Raji-luciferase cells and a reduced number of CAR-T cells, and observed similar results (Figures S4H–S4K). Since all these experiments are based on T cells from a single human donor, we evaluated anti-tumor capability of MAG-CD19 relative to LV-CD19 using T cells from another donor (Figure S4L; STAR Methods). Again, MAG-CD19 CAR-T cells showed enhanced cytotoxicity (Figures S4M and S4N) and superior *in vivo* efficacy (Figures S4O and S4P). The robust efficacy of MAG-CD19 was consistent with its persistence in the peripheral blood of treated mice (Figure S4Q).

Analogous *in vitro* and *in vivo* assays demonstrated the superior efficacy of MAG-CAR-T cells in the treatment of solid tumors relative to LV- and SB-based approaches. Three types of anti-HER1 CAR-T cells were engineered: LV-HER1, SB-HER1, and MAG-HER1 (STAR Methods). In contrast to the anti-CD19 scenario (Figures S4A–S4C), FUW now displays only moderately higher activity compared with pRRLSIN in the multiple rounds of tumor challenge assay (Figures S5A–S5C). Consequently,

we performed evaluations against each of these two LV vectors. In FUW-based assays, despite the high proportion of CAR-positive cells for FUW (Figure 6I), MAG-HER1 cells demonstrated high resistance during *in vitro* challenges, particularly evident in the final round (Figure 6J). Consistently, *in vivo* mouse data indicated that MAG-HER1 CAR-T cells promptly controlled tumor growth (Figures 6K and 6L), resulting in smaller tumors compared with the LV and SB groups (Figures 6M and 6N). Subsequently, for mice showing complete tumor clearance in the SB and MAG groups, re-inoculated tumors were swiftly eliminated (Figure 6O), aligning with the persistent presence of CAR-T cells in the peripheral blood of mice in these two groups (Figure S5D). Evaluation results based on pRRLSIN largely recapitulated the superiority of MAG-HER1 CAR-T cells (Figures S5E–S5J), including high *in vitro* resistance of tumor cells (Figures S5F and S5G), rapid *in vivo* tumor elimination (Figures S5H and S5I), and sustained presence in the blood (Figure S5J).

Given potentially distinct functionalities across CARs with different co-stimulatory domains,⁹⁶ we substituted the widely used 4-1BB domain utilized in previous experiments with the CD28 domain (Figures 6B, S4A, and S5A). MAG-CD19/HER1 CAR-T cells again surpassed LV and SB CAR-T cells (Figures S4R, S4S, S5K, and S5L).

We then investigated the mechanisms underlying the performance of TE-based CAR-T cells, particularly MAG cells. Given similar T cell composition across different CAR-T cells (Figure S4F), the superiority of TE-based cells likely stems from their persistence (e.g., Figures S4G and S5G). Notably, a main challenge to sustained CAR functionality is the potential for CARs to elevate tonic signaling prior to tumor cell exposure, leading to susceptibility to activation-induced cell death (AICD) and premature T cell exhaustion.^{97,98} To explore this possibility, we exposed cells to continuous antigen stimulation, establishing a model mimicking CAR-T cell exhaustion for both anti-CD19 and anti-HER1 therapies (Figures S6A and S6B; STAR Methods). Before stimulation, quantification of specific markers revealed comparable levels of tonic signaling and AICD across LV, SB, and MAG CAR-T cells, with a slight elevation of phospho-CD3 ζ (a tonic signaling marker) in MAG cells and Fas (an AICD marker) in LV cells (Figures S6C and S6D). Similarly, all three cell types

Figure 6. Anti-tumor capability of lentiviral and TE-derived CAR-T cells

- (A) The transposition activity of TEs in human T cells. Data are represented as mean $\pm$ SD across two biological replicates.
- (B) The workflow of lentiviral and TE-derived CAR-T cell generation and functional evaluation. EF1 α and CMV promoters were used. For additional abbreviations, please refer to the legend of Figure S4A.
- (C) CAR expression quantification by flow cytometry analysis of CD19 CAR-T cells on the 12th day after electroporation. The terms "Freq" and "MFI" (mean fluorescence intensity) denote the percentage of CAR-positive cells and the expression level of CAR, respectively.
- (D) Tumor cell lysis ability of CD19 CAR-T cells across four rounds of incubation with K562-CD19 cells. Data are represented as mean $\pm$ SEM ($n = 3$).
- (E) Schematic of the *in vivo* experimental design using cell line-derived xenograft (CDX) models for CD19 CAR-T cells.
- (F) Bioluminescence imaging results for five mice inoculated with Raji-luciferase tumor cells (Raji-luciferase).
- (G) Quantification of luminescence. The average radiance (p.s./cm/s) represents the tumor burden. Data are represented as mean $\pm$ SEM ($n = 5$).
- (H) The survival curve of tumor-bearing mice.
- (I) CAR expression quantification by flow cytometry analysis of HER1 CAR-T cells on the 12th day post-electroporation.
- (J) Tumor cell lysis ability of HER1 CAR-T cells across five rounds of incubation with H266 cell lines. Data are represented as mean $\pm$ SEM ($n = 3$).
- (K) Schematic of the *in vivo* experimental design using CDX models for HER1 CAR-T cells.
- (L) Tumor volume of mice bearing HER1⁺ tumors. Data are represented as mean $\pm$ SEM ($n = 5$).
- (M) Tumor status in mice on the 52nd day after HER1 CAR-T cell injection. "NA" denotes complete tumor elimination. Data are represented as mean $\pm$ SEM ($n = 5$).
- (N) Tumor weight in mice on the 52nd day after HER1 CAR-T cell injection. Data are represented as mean $\pm$ SEM ($n = 5$).
- (O) Tumor volume in mice after tumor re-inoculation. Data are represented as mean $\pm$ SEM ($n = 5$).

See also Figures S4, S5, and S6.

showed analogous levels of exhaustion in the early stages of stimulation, as evidenced by the expression of four marker genes (Figure S6E). However, LV cells exhibited significantly heightened exhaustion in later stages. Thus, although the extent of antigen-independent tonic signaling and AICD is similar across LV, SB, and MAG CAR-T cells, LV CAR-T cells demonstrate earlier exhaustion following antigen stimulation. Considering the importance of sustained CAR expression for CAR-T cell function^{99–101} and the moderate insertion bias of SB and MAG toward transcriptionally active regions in K562 cells (Figure 5A), it is conceivable that this bias persists in exhausted T cells, contributing to their sustained functionality. Indeed, analyses of insertion site mapping data in T cells and public chromatin accessibility data from a T cell exhaustion model revealed that approximately 9.1% of MAG insertions were located in open chromatin pre-stimulation, with 81% of them remaining open in dysfunctional or exhausted T cells (Figure S6F). In comparison, these percentages were 3.8% and 73% for SB and 2.3% and 77% for LV, respectively (Figure S6G). Thus, the TE insertion bias could be important in prolonging the CAR expression and functionality of TE-based CAR-T cells, especially MAG cells.

Taken together, our analyses of hematological and solid tumors demonstrated the superiority of MAG CAR-T cells over LV CAR-T cells in both *in vitro* and *in vivo* contexts. Notably, MAG CAR-T cells surpassed SB CAR-T cells in most comparisons, particularly in the *in vivo* evaluations.

DISCUSSION

By screening the largest set of DNA TEs in human cells to date, we not only provide insights into the factors underlying the transposition activity and evolutionary dynamics but also expand the TE-based genome engineering toolbox.

First, based on the unified activity dataset, including 130 DNA TEs, we dissected evolutionary and functional factors associated with transposition activity. From an evolutionary standpoint, four of five superfamilies consist of members active in human cells, and the Tc1/*mariner* superfamily, especially the Tc1 family, is enriched with high-activity members (Figures 2A, 2B, and 2D). Furthermore, possibly because the number of autonomous copies directly reflect the recent mobilization of TE, it best predicts the activity of TEs (Figure 2G). From a functional standpoint, the general essentiality of TIRs has been corroborated (Figure 3F), while Tc1 elements with the IR-DR structure tend to be active (Figures 2A and 3A). In addition to TIRs, active TEs harbor critical amino acids, including known functional motifs and their proximal residues (Figures 4A and 4E). All these features such as high copy number or essential residues are valuable across multiple aspects. Specifically, although our computational framework (Figure 1A) demonstrates decent performance, with 30% of candidates being validated as active (Figure 1E), it could be further enhanced by incorporating copy number or residue information. Moreover, by introducing back-mutation into MAG at 11 sites shared across most active TEs but altered in MAG (Figure 4A), our pilot study identified one mutation (M235Q) enhancing the activity by ~30% (Figures S7A and S7B; STAR Methods). Presumably, other novel active TEs could be similarly improved. Essential residue data can also be em-

ployed to evaluate widely used tools such as PolyPhen-2¹⁰² and ESM1b.¹⁰³ Although they effectively predict critical residues within Tc1 elements (Figure S7C; STAR Methods), their performance diminishes for hAT elements (Figure S7D), indicating the necessity for optimization across protein families.

Second, our analyses substantiate horizontal transfer, vertical inactivation, and stochastic loss model.^{20–22} Tc1/*mariner* elements are widely recognized for their susceptibility to horizontal transfer (Figure 2E).^{44–46,104} Our cross-superfamily survey now provides a mechanistic basis, revealing that the Tc1/*mariner* superfamily tends to show high heterologous activity (Figure 2D). Furthermore, whether MITEs are generally associated with various TE families during vertical inactivation is less clear. Our analysis demonstrates that this is not always the case, especially for the Tc1 family, in which MITEs are either absent or only contribute a proportion of non-autonomous TEs (Figures 3E and S2A). This peculiarity may be attributed to the small size of Tc1 elements compared with other TEs (median: 1.6 kb vs. 2.4–5.8 kb, Figure S2I), suggesting that size reduction due to internal deletion may not confer a competitive binding capability for Tc1 MITEs.

Third, our screening has expanded the DNA TE-based genome engineering toolbox, especially by incorporating highly potent MAG. Case studies in the last three decades cumulatively identified 20 DNA TEs active in mammalian cells (Table S1). In this study, we identified 40 cases and increased phylogenetic diversity (Figure 2A). Moreover, our head-to-head characterization revealed the functional divergence of TEs (Figure 5). These data could guide future TE-based applications in either mutagenesis or transgenesis across a wide range of species, given the activity of TEs in both diverse host species and human (Table S2). For mutagenesis, since the TE toolbox consists of semi-randomly inserted members or members preferring transcriptionally active or inactive regions (Figure 5A), multiple members with complementary biases target a genome more comprehensively, which would be valuable for enhancer or gene mapping, as well as random insertion.^{74,85,105} For transgenesis, our CAR-T therapy experiments (Figures 6D–6H and 6J–6O) showed that MAG largely outperformed the conventionally used lentivirus and SB100X. The significance of these results is 2-fold. On the one hand, except codon optimization, MAG has not been subjected to engineering, while SB100X went through five generations of optimization over 10 years.³⁸ Nonetheless, MAG largely outcompeted SB100X in terms of the activity and CAR-T therapy performance (Figures 1E and 6), highlighting the performance of the data-driven strategy for tool development. Note that MAG could be further enhanced based on optimization strategies developed for SB. For instance, among the five known SB mutations⁸³ that reduce genic integration bias (Figure 5A) and consequently lower the risk of oncogenic transformation, three exhibited a similar effect for MAG without heavily impacting its activity (Figures S7E–S7G). On the other hand, MAG shows superior performance in CAR-T therapy against both hematological and solid tumors, while DNA TE-based CAR-T therapy against solid tumors has been rarely reported. Together with the simplicity and cost-effectiveness of TE-based CAR-T cell generation compared with lentiviral-based approaches,^{41,106,107} MAG represents a potent platform for developing CAR-T therapies targeting

various tumors. In addition, it is important to note the recently increasing popularity of CRISPR-Cas9-mediated knockin techniques in CAR-T therapy (e.g., the insertion of CD19 CAR into the T cell receptor α [TRAC] locus^{108,109}). This approach offers the advantage of uniform endogenous regulation of CAR expression. Despite potential drawbacks such as undesired editing outcomes, low integration efficiency, or the lack of long-term follow-up, the knockin technique holds promise and supplements conventional lentiviral or TE-based techniques.^{97,110,111}

In conclusion, our systematic and comparative framework complements conventional case studies in illuminating basic biology and empowering applied biology. Moreover, we have highlighted the idiosyncratic nature of DNA TEs in transposition activity, evolutionary dynamics, and application potential. Shaped by the relentless arms race between hosts and TEs, diverse TEs will continue to fascinate us, as they fascinated McClintock 70 years ago.

Limitations of the study

Despite the substantial number of TEs screened (130), our dataset is undoubtedly incomplete. The insights gleaned from this dataset are biased toward the overrepresented Tc1/*mariner* and *hAT* elements. Future investigations into underrepresented superfamilies or families hold promise for uncovering novel evolutionary or functional insights. Moreover, the 40 newly discovered active TEs are only partially characterized. Even for the extensively studied TE like MAG, follow-up studies are needed to explore additional factors beyond insertion bias influencing its CAR-T therapy potency, to optimize its efficiency and safety, and to compare its performance against alternative CAR-T systems, particularly the TRAC knockin system.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
 - Cell lines
 - Primary human T cells
 - Animals
- METHOD DETAILS
 - Literature survey of DNA TEs active in mammalian cells
 - Identification of potentially active DNA TEs
 - Phylogenetic modeling of activity-related factors
 - Identification of nonautonomous TE copies
 - Classification of TE evolutionary dynamics in terms of MITES
 - Analyses of direct repeats in the Tc1/*mariner* superfamily
 - Identification of critical residues associated with transposition activity
 - Protein structural modeling
 - Plasmid construction
 - Transposition assay
 - Insertion site analyses
 - Cargo size tolerance of the top active TEs in the three superfamilies
 - Overproduction inhibition assays

- CAR-T-related experiments
- A pilot study on optimization of MAG
- Study approval
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cell.2024.05.007>.

ACKNOWLEDGMENTS

We are grateful to Drs. Jinfeng Chen, Xionglei He, Shujun Ou, Lu Wang, Cheng Sun, and Xiaoyue Wang for their insightful comments. This research was supported by the National Natural Science Foundation of China (32325014 to Y.E.Z.), the Ministry of Agriculture and Rural Affairs of China (2023ZD0407401 to H.W.), the National Key R&D Program of China (2019YFA0802600 to Y.E.Z. and 2019YFA0110000 to H.W.), the Chinese Academy of Sciences (ZDBS-LY-SM005 to Y.E.Z.), and the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA16010503 to H.W.).

AUTHOR CONTRIBUTIONS

Y.E.Z., H.W., T.Z., and S.T. conceived and designed the study. T.Z. and N.T. performed experiments with the help of J.S., H.G., C.Z., C.W., W.S., X.Z., L.F., and Y.W. P.G. and W.Y. provided the technical support. S.T. and Y.L. performed computational analyses with the help of T.Z., Y.G., H.M., X.H., Q.C., Y.Z., and B.G. Y.E.Z., H.W., T.Z., S.T., Y.L., and N.T. wrote the manuscript.

DECLARATION OF INTERESTS

A patent application (H.W., Y.Z., S.T., and T.Z. as inventors) has been filed by the Institute of Zoology, Chinese Academy of Sciences.

Received: August 7, 2023

Revised: March 11, 2024

Accepted: May 2, 2024

Published: June 5, 2024

REFERENCES

1. McClintock, B. (1948). Mutable loci in maize. *Carnegie Institute of Washington Year Book* 47, 155–169.
2. McClintock, B. (1950). The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci. USA* 36, 344–355. <https://doi.org/10.1073/pnas.36.6.344>.
3. McClintock, B. (1956). Controlling elements and the gene. *Cold Spring Harb. Symp. Quant. Biol.* 21, 197–216. <https://doi.org/10.1101/sqb.1956.021.01.017>.
4. Barrón, M.G., Fiston-Lavier, A.-S., Petrov, D.A., and González, J. (2014). Population genomics of transposable elements in *Drosophila*. *Annu. Rev. Genet.* 48, 561–581. <https://doi.org/10.1146/annurev-genet-120213-092359>.
5. Wells, J.N., and Feschotte, C. (2020). A field guide to eukaryotic transposable elements. *Annu. Rev. Genet.* 54, 539–561. <https://doi.org/10.1146/annurev-genet-040620-022145>.
6. Feschotte, C., Jiang, N., and Wessler, S.R. (2002). Plant transposable elements: where genetics meets genomics. *Nat. Rev. Genet.* 3, 329–341. <https://doi.org/10.1038/nrg793>.
7. Cosby, R.L., Judd, J., Zhang, R., Zhong, A., Garry, N., Pritham, E.J., and Feschotte, C. (2021). Recurrent evolution of vertebrate transcription factors by transposase capture. *Science* 371, eabc6405. <https://doi.org/10.1126/science.abc6405>.

8. Jangam, D., Feschotte, C., and Betrán, E. (2017). Transposable element domestication as an adaptation to evolutionary conflicts. *Trends Genet.* 33, 817–831. <https://doi.org/10.1016/j.tig.2017.07.011>.
9. Rech, G.E., Radó, S., Guirao-Rico, S., Aguilera, L., Horvath, V., Green, L., Lindstadt, H., Jamilloux, V., Quesneville, H., and González, J. (2022). Population-scale long-read sequencing uncovers transposable elements associated with gene expression variation and adaptive signatures in *Drosophila*. *Nat. Commun.* 13, 1948. <https://doi.org/10.1038/s41467-022-29518-8>.
10. Almeida, M.V., Vernaz, G., Putman, A.L.K., and Miska, E.A. (2022). Taming transposable elements in vertebrates: from epigenetic silencing to domestication. *Trends Genet.* 38, 529–553. <https://doi.org/10.1016/j.tig.2022.02.009>.
11. Plasterk, R.H., Izsvák, Z., and Ivics, Z. (1999). Resident aliens: the Tc1/mariner superfamily of transposable elements. *Trends Genet.* 15, 326–332. [https://doi.org/10.1016/s0168-9525\(99\)01777-1](https://doi.org/10.1016/s0168-9525(99)01777-1).
12. Yuan, Y.W., and Wessler, S.R. (2011). The catalytic domain of all eukaryotic cut-and-paste transposase superfamilies. *Proc. Natl. Acad. Sci. USA* 108, 7884–7889. <https://doi.org/10.1073/pnas.1104208108>.
13. Liu, K., and Wessler, S.R. (2017). Functional characterization of the active Mutator-like transposable element, Muta1 from the mosquito *Aedes aegypti*. *Mobile DNA* 8, 1. <https://doi.org/10.1186/s13100-016-0084-6>.
14. Richardson, J.M., Colloms, S.D., Finnegan, D.J., and Walkinshaw, M.D. (2009). Molecular architecture of the Mos1 paired-end complex: the structural basis of DNA transposition in a eukaryote. *Cell* 138, 1096–1108. <https://doi.org/10.1016/j.cell.2009.07.012>.
15. Atkinson, P.W. (2015). hAT transposable elements. *Microbiol. Spec.* 3, 773–800. <https://doi.org/10.1128/microbiolspec.MDNA3-0054-2014>.
16. Morellet, N., Li, X., Wieninger, S.A., Taylor, J.L., Bischerour, J., Moriau, S., Lescop, E., Bardiaux, B., Mathy, N., and Assrir, N. (2018). Sequence-specific DNA binding activity of the cross-brace zinc finger motif of the piggyBac transposase. *Nucleic Acids Res.* 46, 2660–2677. <https://doi.org/10.1093/nar/gky044>.
17. Voigt, F., Wiedemann, L., Zuliani, C., Querques, I., Sebe, A., Mátés, L., Izsvák, Z., Ivics, Z., and Barabas, O. (2016). Sleeping Beauty transposase structure allows rational design of hyperactive variants for genetic engineering. *Nat. Commun.* 7, 11126. <https://doi.org/10.1038/ncomms11126>.
18. Cui, Z., Geurts, A.M., Liu, G., Kaufman, C.D., and Hackett, P.B. (2002). Structure–function analysis of the inverted terminal repeats of the Sleeping Beauty transposon. *J. Mol. Biol.* 318, 1221–1235. [https://doi.org/10.1016/s0022-2836\(02\)00237-1](https://doi.org/10.1016/s0022-2836(02)00237-1).
19. Izsvák, Z., Ivics, Z., and Plasterk, R.H. (2000). Sleeping Beauty, a wide host-range transposon vector for genetic transformation in vertebrates. *J. Mol. Biol.* 302, 93–102. <https://doi.org/10.1006/jmbi.2000.4047>.
20. Feschotte, C., and Pritham, E.J. (2007). DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* 41, 331–368. <https://doi.org/10.1146/annurev.genet.40.110405.090448>.
21. Lohe, A.R., Moriyama, E.N., Lidholm, D.-A., and Hartl, D.L. (1995). Horizontal transmission, vertical inactivation, and stochastic loss of mariner-like transposable elements. *Mol. Biol. Evol.* 12, 62–72. <https://doi.org/10.1093/oxfordjournals.molbev.a040191>.
22. Robillard, É., Le Rouzic, A., Zhang, Z., Capi, P., and Hua-Van, A. (2016). Experimental evolution reveals hyperparasitic interactions among transposable elements. *Proc. Natl. Acad. Sci. USA* 113, 14763–14768. <https://doi.org/10.1073/pnas.1524143113>.
23. Moon, S., Cassani, M., Lin, Y.A., Wang, L., Dou, K., and Zhang, Z.Z. (2018). A robust transposon–endogenizing response from germline stem cells. *Dev. Cell* 47, 660–671.e3. <https://doi.org/10.1016/j.devcel.2018.10.011>.
24. Chen, J., Lu, L., Robb, S.M.C., Collin, M., Okumoto, Y., Stajich, J.E., and Wessler, S.R. (2020). Genomic diversity generated by a transposable element burst in a rice recombinant inbred population. *Proc. Natl. Acad. Sci. USA* 117, 26288–26297. <https://doi.org/10.1073/pnas.2015736117>.
25. Zhao, D., Ferguson, A.A., and Jiang, N. (2016). What makes up plant genomes: the vanishing line between transposable elements and genes. *Biochim. Biophys. Acta* 1859, 366–380. <https://doi.org/10.1016/j.bbagr.2015.12.005>.
26. Pace, J.K., Gilbert, C., Clark, M.S., and Feschotte, C. (2008). Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods. *Proc. Natl. Acad. Sci. USA* 105, 17023–17028. <https://doi.org/10.1073/pnas.0806548105>.
27. Feschotte, C., and Mouchès, C. (2000). Evidence that a family of miniature inverted-repeat transposable elements (MITs) from the *Arabidopsis thaliana* genome has arisen from a pogo-like DNA transposon. *Mol. Biol. Evol.* 17, 730–737. <https://doi.org/10.1093/oxfordjournals.molbev.a026351>.
28. Zhao, D., Ferguson, A., and Jiang, N. (2015). Transposition of a rice Mutator-like element in the yeast *Saccharomyces cerevisiae*. *Plant Cell* 27, 132–148. <https://doi.org/10.1105/tpc.114.128488>.
29. Feschotte, C., Zhang, X., and Wessler, S.R. (2007). Miniature inverted-repeat transposable elements and their relationship to established DNA transposons. *Genome* 56, 475–486. <https://doi.org/10.1139/gen-2012-0174>.
30. Rubin, G.M., and Spradling, A.C. (1982). Genetic transformation of *Drosophila* with transposable element vectors. *Science* 218, 348–353. <https://doi.org/10.1126/science.6289436>.
31. Ivics, Z., Hackett, P.B., Plasterk, R.H., and Izsvák, Z. (1997). Z.I. *Cell* 91, 501–510. [https://doi.org/10.1016/s0092-8674\(00\)80436-5](https://doi.org/10.1016/s0092-8674(00)80436-5).
32. Yusa, K., Zhou, L., Li, M.A., Bradley, A., and Craig, N.L. (2011). A hyperactive piggyBac transposase for mammalian applications. *Proc. Natl. Acad. Sci. USA* 108, 1531–1536. <https://doi.org/10.1073/pnas.1008322108>.
33. Sinzelle, L., Kapitonov, V.V., Grzela, D.P., Jursch, T., Jurka, J., Izsvák, Z., and Ivics, Z. (2008). Transposition of a reconstructed Harbinger element in human cells and functional homology with two transposon-derived cellular genes. *Proc. Natl. Acad. Sci. USA* 105, 4715–4720. <https://doi.org/10.1073/pnas.0707746105>.
34. Wang, S., Gao, B., Miskey, C., Guan, Z., Sang, Y., Chen, C., Wang, X., Ivics, Z., and Song, C. (2023). Passer, a highly active transposon from a fish genome, as a potential new robust genetic manipulation tool. *Nucleic Acids Res.* 51, 1843–1858. <https://doi.org/10.1093/nar/gkad005>.
35. Moudgil, A., Wilkinson, M.N., Chen, X., He, J., Cammack, A.J., Vasek, M.J., Lagunas, T., Jr., Qi, Z., Lalli, M.A., Guo, C., et al. (2020). Self-reporting transposons enable simultaneous readout of gene expression and transcription factor binding in single cells. *Cell* 182, 992–1008.e21. <https://doi.org/10.1016/j.cell.2020.06.037>.
36. Palazzo, A., Caizzi, R., Moschetti, R., and Marsano, R.M. (2022). What have we learned in 30 years of investigations on Bari transposons? *Cells* 11, 583. <https://doi.org/10.3390/cells11030583>.
37. Qi, Z., Wilkinson, M.N., Chen, X., Sankaraman, S., Mayhew, D., and Mitra, R.D. (2017). An optimized, broadly applicable piggyBac transposon induction system. *Nucleic Acids Res.* 45, e55. <https://doi.org/10.1093/nar/gkw1290>.
38. Sandoval-Villegas, N., Nurieva, W., Amberger, M., and Ivics, Z. (2021). Contemporary transposon tools: a review and guide through mechanisms and applications of sleeping beauty, piggyBac and Tol2 for genome engineering. *Int. J. Mol. Sci.* 22, 5084. <https://doi.org/10.3390/ijms22105084>.
39. Kebriaei, P., Izsvák, Z., Narayanavari, S.A., Singh, H., and Ivics, Z. (2017). Gene therapy with the sleeping beauty transposon system. *Trends Genet.* 33, 852–870. <https://doi.org/10.1016/j.tig.2017.08.008>.
40. Amberger, M., and Ivics, Z. (2020). Latest advances for the sleeping beauty transposon system: 23 years of insomnia but prettier than ever: refinement and recent innovations of the sleeping beauty transposon

- p>system enabling novel, nonviral genetic engineering applications.
- Bio-Essays*
- 42, e2000136.
- <https://doi.org/10.1002/bies.202000136>
- .
41. Moretti, A., Ponzo, M., Nicolette, C.A., Tcherepanova, I.Y., Biondi, A., and Magnani, C.F. (2022). The past, present, and future of non-viral CAR T cells. *Front. Immunol.* 13, 867013. <https://doi.org/10.3389/fimmu.2022.867013>.
 42. Smit, A.F., and Riggs, A.D. (1996). Tiggers and DNA transposon fossils in the human genome. *Proc. Natl. Acad. Sci. USA* 93, 1443–1448. <https://doi.org/10.1073/pnas.93.4.1443>.
 43. Ivics, Z., Li, M.A., Mátés, L., Boeke, J.D., Nagy, A., Bradley, A., and Izsvák, Z. (2009). Transposon-mediated genome manipulation in vertebrates. *Nat. Methods* 6, 415–422. <https://doi.org/10.1038/nmeth.1332>.
 44. Reiss, D., Mialdea, G., Miele, V., de Vienne, D.M., Peccoud, J., Gilbert, C., Duret, L., and Charlat, S. (2019). Global survey of mobile DNA horizontal transfer in arthropods reveals Lepidoptera as a prime hotspot. *PLoS Genet.* 15, e1007965. <https://doi.org/10.1371/journal.pgen.1007965>.
 45. Zhang, H.-H., Peccoud, J., Xu, M.-R.-X., Zhang, X.-G., and Gilbert, C. (2020). Horizontal transfer and evolution of transposable elements in vertebrates. *Nat. Commun.* 11, 1362. <https://doi.org/10.1038/s41467-020-15149-4>.
 46. Gilbert, C., and Feschotte, C. (2018). Horizontal acquisition of transposable elements and viral sequences: patterns and consequences. *Curr. Opin. Genet. Dev.* 49, 15–24. <https://doi.org/10.1016/j.gde.2018.02.007>.
 47. Dorman, J., Grey, H., and Richardson, J.M. (2015). Structural role of the flanking DNA in mariner transposon excision. *Nucleic Acids Res.* 43, 2424–2432. <https://doi.org/10.1093/nar/gkv096>.
 48. Rhead, B., Karolchik, D., Kuhn, R.M., Hinrichs, A.S., Zweig, A.S., Fujita, P.A., Diekhans, M., Smith, K.E., Rosenbloom, K.R., Raney, B.J., et al. (2010). The UCSC Genome Browser database: update 2010. *Nucleic Acids Res.* 38, D613–D619. <https://doi.org/10.1093/nar/gkp939>.
 49. Bao, W., Kojima, K.K., and Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA* 6, 11. <https://doi.org/10.1186/s13100-015-0041-9>.
 50. Horváth, V., Merenciano, M., and González, J. (2017). Revisiting the relationship between transposable elements and the eukaryotic stress response. *Trends Genet.* 33, 832–841. <https://doi.org/10.1016/j.tig.2017.08.007>.
 51. Ray, D.A., Feschotte, C., Pagan, H.J.T., Smith, J.D., Pritham, E.J., Arensburger, P., Atkinson, P.W., and Craig, N.L. (2008). Multiple waves of recent DNA transposon activity in the bat, *Myotis lucifugus*. *Genome Res.* 18, 717–728. <https://doi.org/10.1101/gr.071886.107>.
 52. Pace, J.K.I., and Feschotte, C. (2007). The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. *Genome Res.* 17, 422–432. <https://doi.org/10.1101/gr.5826307>.
 53. Hellsten, U., Harland, R.M., Gilchrist, M.J., Hendrix, D., Jurka, J., Kapitonov, V., Ovcharenko, I., Putnam, N.H., Shu, S., Taher, L., et al. (2010). The genome of the Western clawed frog *Xenopus tropicalis*. *Science* 328, 633–636. <https://doi.org/10.1126/science.1183670>.
 54. Tan, S., Ma, H., Wang, J., Wang, M., Wang, M., Yin, H., Zhang, Y., Zhang, X., Shen, J., Wang, D., et al. (2021). DNA transposons mediate duplications via transposition-independent and -dependent mechanisms in metazoans. *Nat. Commun.* 12, 4280. <https://doi.org/10.1038/s41467-021-24585-9>.
 55. Almojil, D., Bourgeois, Y., Falis, M., Hariyani, I., Wilcox, J., and Boissinot, S. (2021). The structural, functional and evolutionary impact of transposable elements in eukaryotes. *Genes* 12, 918. <https://doi.org/10.3390/genes12060918>.
 56. Majumdar, S., and Rio, D.C. (2015). P transposable elements in *Drosophila* and other eukaryotic organisms. *Microbiol. Spec.* 3, MDNA3–MDNA4. <https://doi.org/10.1128/microbiolspec.MDNA3-0004-2014>.
 57. Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., et al. (2007). A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* 8, 973–982. <https://doi.org/10.1038/nrg2165>.
 58. Ding, S., Wu, X., Li, G., Han, M., Zhuang, Y., and Xu, T. (2005). Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. *Cell* 122, 473–483. <https://doi.org/10.1016/j.cell.2005.07.013>.
 59. Shen, D., Song, C., Miskey, C., Chan, S., Guan, Z., Sang, Y., Wang, Y., Chen, C., Wang, X., Müller, F., et al. (2021). A native, highly active Tc1/mariner transposon from zebrafish (ZB) offers an efficient genetic manipulation tool for vertebrates. *Nucleic Acids Res.* 49, 2126–2140. <https://doi.org/10.1093/nar/gkab045>.
 60. Guan, Z., Shi, S., Diaby, M., Danley, P., Ullah, N., Puzakov, M., Gao, B., and Song, C. (2022). Horizontal transfer of Buster transposons across multiple phyla and classes of animals. *Mol. Phylogenet. Evol.* 173, 107506. <https://doi.org/10.1016/j.ympev.2022.107506>.
 61. Arensburger, P., Hice, R.H., Zhou, L., Smith, R.C., Tom, A.C., Wright, J.A., Knapp, J., O'Brochta, D.A., Craig, N.L., and Atkinson, P.W. (2011). Phylogenetic and functional characterization of the hAT transposon superfamily. *Genetics* 188, 45–57. <https://doi.org/10.1534/genetics.111.126813>.
 62. Habu, Y., Hisatomi, Y., and Iida, S. (1998). Molecular characterization of the mutable flaked allele for flower variegation in the common morning glory. *Plant J.* 16, 371–376. <https://doi.org/10.1046/j.1365-3113x.1998.00308.x>.
 63. Rossato, D.O., Ludwig, A., Deprá, M., Loreto, E.L.S., Ruiz, A., and Valente, V.L.S. (2014). BuT2 is a member of the third major group of hAT transposons and is involved in horizontal transfer events in the genus *Drosophila*. *Genome Biol. Evol.* 6, 352–365. <https://doi.org/10.1093/gbe/evu017>.
 64. Peccoud, J., Loiseau, V., Cordaux, R., and Gilbert, C. (2017). Massive horizontal transfer of transposable elements in insects. *Proc. Natl. Acad. Sci. USA* 114, 4721–4726. <https://doi.org/10.1073/pnas.1621178114>.
 65. Hartl, D.L., Lohe, A.R., and Lozovskaya, E.R. (1997). Regulation of the transposable element mariner. *Genetica* 100, 177–184. https://doi.org/10.1007/978-94-011-4898-6_18.
 66. O'Brochta, D.A., and Handler, A.M. (1988). Mobility of P elements in drosophilids and nondrosophilids. *Proc. Natl. Acad. Sci. USA* 85, 6052–6056. <https://doi.org/10.1073/pnas.85.16.6052>.
 67. Silva, J.C., Loreto, E.L., and Clark, J.B. (2004). Factors that affect the horizontal transfer of transposable elements. *Curr. Issues Mol. Biol.* 6, 57–71.
 68. Moschetti, R., Chlamydas, S., Marsano, R.M., and Caizzi, R. (2008). Conserved motifs and dynamic aspects of the terminal inverted repeat organization within Bari-like transposons. *Mol. Genet. Genomics* 279, 451–461. <https://doi.org/10.1007/s00438-008-0324-7>.
 69. Ivics, Z., Izsvák, Z., Minter, A., and Hackett, P.B. (1996). Identification of functional domains and evolution of Tc1-like transposable elements. *Proc. Natl. Acad. Sci. USA* 93, 5008–5013. <https://doi.org/10.1073/pnas.93.10.5008>.
 70. Gao, B., Wang, Y., Diaby, M., Zong, W., Shen, D., Wang, S., Chen, C., Wang, X., and Song, C. (2020). Evolution of pogo, a separate superfamily of IS630-Tc1-mariner transposons, revealing recurrent domestication events in vertebrates. *Mobile DNA* 11, 25. <https://doi.org/10.1186/s13100-020-00220-0>.
 71. Yang, G., Nagel, D.H., Feschotte, C., Hancock, C.N., and Wessler, S.R. (2009). Tuned for transposition: molecular determinants underlying the hyperactivity of a Stowaway MITE. *Science* 325, 1391–1394. <https://doi.org/10.1126/science.1175688>.
 72. Hickman, A.B., Ewis, H.E., Li, X., Knapp, J.A., Laver, T., Doss, A.-L., Tolun, G., Steven, A.C., Grishaev, A., and Bax, A. (2014). Structural basis of hAT transposon end recognition by Hermes, an octameric DNA transposase from *Musca domestica*. *Cell* 158, 353–367. <https://doi.org/10.1016/j.cell.2014.05.037>.

73. Tosi, L.R., and Beverley, S.M. (2000). cis and trans factors affecting Mos1 mariner evolution and transposition in vitro, and its potential for functional genomics. *Nucleic Acids Res.* 28, 784–790. <https://doi.org/10.1093/nar/28.3.784>.
74. Pavlopoulos, A., Oehler, S., Kapetanaki, M.G., and Savakis, C. (2007). The DNA transposon Minos as a tool for transgenesis and functional genomic analysis in vertebrates and invertebrates. *Genome Biol.* 8, S2. <https://doi.org/10.1186/gb-2007-8-s1-s2>.
75. Lampe, D.J., Churchill, M.E., and Robertson, H.M. (1996). A purified mariner transposase is sufficient to mediate transposition in vitro. *EMBO J.* 15, 5470–5479. <https://doi.org/10.1002/j.1460-2075.1996.tb00930.x>.
76. Tellier, M., Bouuaert, C.C., and Chalmers, R. (2015). Mariner and the ITm superfamily of transposons. *Microbiol. Spectr.* 3, MDNA3–MDN33. <https://doi.org/10.1128/microbiolspec.MDNA3-0033-2014>.
77. Morris, E.R., Grey, H., McKenzie, G., Jones, A.C., and Richardson, J.M. (2016). A bend, flip and trap mechanism for transposon integration. *eLife* 5, e15537. <https://doi.org/10.7554/eLife.15537>.
78. Grabundzija, I., Irgang, M., Mátés, L., Belay, E., Matrai, J., Gogol-Döring, A., Kawakami, K., Chen, W., Ruiz, P., and Chuah, M.K. (2010). Comparative analysis of transposable element vector systems in human cells. *Mol. Ther.* 18, 1200–1209. <https://doi.org/10.1038/mt.2010.47>.
79. Roadmap; Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenyk, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., et al. (2015). Integrative analysis of 111 reference human epigenomes. *Nature* 518, 317–330. <https://doi.org/10.1038/nature14248>.
80. Craig, N.L. (2015). A moveable feast: an introduction to mobile DNA. *Mobile DNA III*, 1–39.
81. Sadelain, M., Papapetrou, E.P., and Bushman, F.D. (2011). Safe harbours for the integration of new DNA in the human genome. *Nat. Rev. Cancer* 12, 51–58. <https://doi.org/10.1038/nrc3179>.
82. Levin, H.L., and Moran, J.V. (2011). Dynamic interactions between transposable elements and their hosts. *Nat. Rev. Genet.* 12, 615–627. <https://doi.org/10.1038/nrg3030>.
83. Miskey, C., Kesselring, L., Querques, I., Abrusán, G., Barabas, O., and Ivics, Z. (2022). Engineered sleeping beauty transposase redirects transposon integration away from genes. *Nucleic Acids Res.* 50, 2807–2825. <https://doi.org/10.1093/nar/gkac092>.
84. Sultana, T., Zamborlini, A., Cristofari, G., and Lesage, P. (2017). Integration site selection by retroviruses and transposable elements in eukaryotes. *Nat. Rev. Genet.* 18, 292–308. <https://doi.org/10.1038/nrg.2017.7>.
85. Woodard, L.E., Li, X., Malani, N., Kaja, A., Hice, R.H., Atkinson, P.W., Bushman, F.D., Craig, N.L., and Wilson, M.H. (2012). Comparative analysis of the recently discovered hAT transposon TcBuster in human cells. *PLoS One* 7, e42666. <https://doi.org/10.1371/journal.pone.0042666>.
86. Li, X., Ewis, H., Hice, R.H., Malani, N., Parker, N., Zhou, L., Feschotte, C., Bushman, F.D., Atkinson, P.W., and Craig, N.L. (2013). A resurrected mammalian hAT transposable element and a closely related insect element are highly active in human cell culture. *Proc. Natl. Acad. Sci. USA* 110, E478–E487. <https://doi.org/10.1073/pnas.1121543109>.
87. Linheiro, R.S., and Bergman, C.M. (2012). Whole genome resequencing reveals natural target site preferences of transposable elements in *Drosophila melanogaster*. *PLoS One* 7, e30008. <https://doi.org/10.1371/journal.pone.0030008>.
88. Kunze, R., Behrens, U., Courage-Franzkowiak, U., Feldmar, S., Kühn, S., and Lütticke, R. (1993). Dominant transposition-deficient mutants of maize Activator (Ac) transposase. *Proc. Natl. Acad. Sci. USA* 90, 7094–7098. <https://doi.org/10.1073/pnas.90.15.7094>.
89. Woodard, L.E., Downes, L.M., Lee, Y.-C., Kaja, A., Terefe, E.S., and Wilson, M.H. (2017). Temporal self-regulation of transposition through host-independent transposase rodlet formation. *Nucleic Acids Res.* 45, 353–366. <https://doi.org/10.1093/nar/gkw115>.
90. Bire, S., Casteret, S., Arnaoty, A., Piégu, B., Lecomte, T., and Bigot, Y. (2013). Transposase concentration controls transposition activity: myth or reality? *Gene* 530, 165–171. <https://doi.org/10.1016/j.gene.2013.08.039>.
91. Lois, C., Hong, E.J., Pease, S., Brown, E.J., and Baltimore, D. (2002). Germline transmission and tissue-specific expression of transgenes delivered by lentiviral vectors. *Science* 295, 868–872. <https://doi.org/10.1126/science.1067081>.
92. Zhang, X., Zhang, C., Qiao, M., Cheng, C., Tang, N., Lu, S., Sun, W., Xu, B., Cao, Y., Wei, X., et al. (2022). Depletion of BATF in CAR-T cells enhances antitumor activity by inducing resistance against exhaustion and formation of central memory cells. *Cancer Cell* 40, 1407–1422.e7. <https://doi.org/10.1016/j.ccell.2022.09.013>.
93. Schmitt, T.M., Aggen, D.H., Ishida-Tsubota, K., Ochsenreither, S., Kranz, D.M., and Greenberg, P.D. (2017). Generation of higher affinity T cell receptors by antigen-driven differentiation of progenitor T cells in vitro. *Nat. Biotechnol.* 35, 1188–1195. <https://doi.org/10.1038/nbt.4004>.
94. Humbert, O., Gisch, D.W., Wohlfahrt, M.E., Adams, A.B., Greenberg, P.D., Schmitt, T.M., Trobridge, G.D., and Kiem, H.-P. (2016). Development of third-generation coccal envelope Producer Cell Lines for Robust Lentiviral Gene Transfer into Hematopoietic Stem Cells and T-cells. *Mol. Ther.* 24, 1237–1246. <https://doi.org/10.1038/mt.2016.70>.
95. Rollins, M.R., Spartz, E.J., and Stromnes, I.M. (2020). T cell receptor engineered lymphocytes for cancer therapy. *Curr. Protoc. Immunol.* 129, e97. <https://doi.org/10.1002/cpim.97>.
96. Cappell, K.M., and Kochenderfer, J.N. (2021). A comparison of chimeric antigen receptors containing CD28 versus 4–1BB costimulatory domains. *Nat. Rev. Clin. Oncol.* 18, 715–727. <https://doi.org/10.1038/s41571-021-00530-z>.
97. Tristán-Manzano, M., Maldonado-Pérez, N., Justicia-Lirio, P., Muñoz, P., Cortijo-Gutiérrez, M., Pavlovic, K., Jiménez-Moreno, R., Nogueras, S., Carmona, M.D., Sánchez-Hernández, S., et al. (2022). Physiological lentiviral vectors for the generation of improved CAR-T cells. *Mol. Ther. Oncolytics* 25, 335–349. <https://doi.org/10.1016/j.omto.2022.05.003>.
98. Rodríguez-Marquez, P., Calleja-Cervantes, M.E., Serrano, G., Oliver-Caldes, A., Palacios-Berraquero, M.L., Martín-Mallo, A., Calviño, C., Español-Rego, M., Ceballos, C., Lozano, T., et al. (2022). CAR density influences antitumoral efficacy of BCMA CAR T cells and correlates with clinical outcome. *Sci. Adv.* 8, eabo0514. <https://doi.org/10.1126/sciadv.abo0514>.
99. Li, W., Qiu, S., Chen, J., Jiang, S., Chen, W., Jiang, J., Wang, F., Si, W., Shu, Y., Wei, P., et al. (2020). Chimeric antigen receptor designed to prevent ubiquitination and downregulation showed durable antitumor efficacy. *Immunity* 53, 456–470.e6. <https://doi.org/10.1016/j.immuni.2020.07.011>.
100. Milone, M.C., Fish, J.D., Carpenito, C., Carroll, R.G., Binder, G.K., Teachey, D., Samanta, M., Lakhal, M., Gloss, B., Danet-Desnoyers, G., et al. (2009). Chimeric receptors containing CD137 signal transduction domains mediate enhanced survival of T cells and increased antileukemic efficacy in vivo. *Mol. Ther.* 17, 1453–1464. <https://doi.org/10.1038/mt.2009.83>.
101. Guedan, S., Posey, A.D., Shaw, C., Wing, A., Da, T., Patel, P.R., McGettigan, S.E., Casado-Medrano, V., Kawalekar, O.U., Uribe-Herranz, M., et al. (2018). Enhancing CAR T cell persistence through ICOS and 4–1BB costimulation. *JCI Insight* 3, e96976. <https://doi.org/10.1172/jci.insight.96976>.
102. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. *Nat. Methods* 7, 248–249. <https://doi.org/10.1038/nmeth0410-248>.
103. Brandes, N., Goldman, G., Wang, C.H., Ye, C.J., and Ntranos, V. (2023). Genome-wide prediction of disease variant effects with a deep protein language model. *Nat. Genet.* 55, 1512–1522. <https://doi.org/10.1038/s41588-023-01465-0>.

104. de Melo, E.S., and Wallau, G.L. (2020). Mosquito genomes are frequently invaded by transposable elements through horizontal transfer. *PLoS Genet.* 16, e1008946. <https://doi.org/10.1371/journal.pgen.1008946>.
105. Kawakami, K., Largaespada, D.A., and Ivics, Z. (2017). Transposons as tools for functional genomics in vertebrate models. *Trends Genet.* 33, 784–801. <https://doi.org/10.1016/j.tig.2017.07.006>.
106. Monjezi, R., Miskey, C., Gogishvili, T., Schleef, M., Schmeer, M., Einsele, H., Ivics, Z., and Hudecek, M. (2017). Enhanced CAR T-cell engineering using non-viral Sleeping Beauty transposition from minicircle vectors. *Leukemia* 31, 186–194. <https://doi.org/10.1038/leu.2016.180>.
107. Hudecek, M., and Ivics, Z. (2018). Non-viral therapeutic cell engineering with the Sleeping Beauty transposon system. *Curr. Opin. Genet. Dev.* 52, 100–108. <https://doi.org/10.1016/j.gde.2018.06.003>.
108. Eyquem, J., Mansilla-Soto, J., Giavridis, T., van der Stegen, S.J.C., Hamieh, M., Cunanan, K.M., Odak, A., Gönen, M., and Sadelain, M. (2017). Targeting a CAR to the TRAC locus with CRISPR/Cas9 enhances tumour rejection. *Nature* 543, 113–117. <https://doi.org/10.1038/nature21405>.
109. An, J., Zhang, C.-P., Qiu, H.-Y., Zhang, H.-X., Chen, Q.-B., Zhang, Y.-M., Lei, X.-L., Zhang, C.-X., Yin, H., and Zhang, Y. (2024). Enhancement of the viability of T cells electroporated with DNA via osmotic dampening of the DNA-sensing cGAS-STING pathway. *Nat. Biomed. Eng.* 8, 149–164. <https://doi.org/10.1038/s41551-023-01073-7>.
110. Nahmad, A.D., Reuveni, E., Goldschmidt, E., Tenne, T., Liberman, M., Horovitz-Fried, M., Khosravi, R., Kobo, H., Reinstein, E., Madi, A., et al. (2022). Frequent aneuploidy in primary human T cells after CRISPR-Cas9 cleavage. *Nat. Biotechnol.* 40, 1807–1813. <https://doi.org/10.1038/s41587-022-01377-0>.
111. Irving, M., Lanitis, E., Migliorini, D., Ivics, Z., and Guedan, S. (2021). Choosing the right tool for genetic engineering: clinical lessons from chimeric antigen receptor-T cells. *Hum. Gene Ther.* 32, 1044–1058. <https://doi.org/10.1089/hum.2021.173>.
112. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. <https://doi.org/10.1038/s41586-021-03819-2>.
113. Paradis, E., Claude, J., and Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20, 289–290. <https://doi.org/10.1093/bioinformatics/btg412>.
114. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. <https://doi.org/10.1093/bioinformatics/btq033>.
115. Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H., and Ferrin, T.E. (2021). UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* 30, 70–82. <https://doi.org/10.1002/pro.3943>.
116. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17, 10–12. <https://doi.org/10.14806/ej.17.1.200>.
117. Hitz, B.C., Lee, J.-W., Jolanki, O., Kagda, M.S., Graham, K., Sud, P., Gabdank, I., Strattan, J.S., Sloan, C.A., Dreszer, T., et al. (2023). The ENCODE uniform analysis pipelines. Preprint at bioRxiv. <https://doi.org/10.1101/2023.04.04.535623>.
118. Burge, C., and Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* 268, 78–94. <https://doi.org/10.1006/jmbi.1997.0951>.
119. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534. <https://doi.org/10.1093/molbev/msaa015>.
120. Letunic, I., and Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. <https://doi.org/10.1093/nar/gkab301>.
121. Yamada, K.D., Tomii, K., and Katoh, K. (2016). Application of the MAFFT sequence alignment program to large data-reexamination of the usefulness of chained guide trees. *Bioinformatics* 32, 3246–3251. <https://doi.org/10.1093/bioinformatics/btw412>.
122. Bailey, T.L., Johnson, J., Grant, C.E., and Noble, W.S. (2015). The MEME suite. *Nucleic Acids Res.* 43, W39–W49. <https://doi.org/10.1093/nar/gkv416>.
123. Madeira, F., Pearce, M., Tivey, A.R.N., Basutkar, P., Lee, J., Edbali, O., Madhusoodanan, N., Kolesnikov, A., and Lopez, R. (2022). Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res.* 50, W276–W279. <https://doi.org/10.1093/nar/gkac240>.
124. Hu, X., Yuan, J., Shi, Y., Lu, J., Liu, B., Li, Z., Chen, Y., Mu, D., Zhang, H., Li, N., et al. (2012). pIRS: profile-based Illumina pair-end reads simulator. *Bioinformatics* 28, 1533–1535. <https://doi.org/10.1093/bioinformatics/bts187>.
125. Smit, A.F.A., Hubley, R., and Green, P. (2013–2015). RepeatMasker Open-4.0. <https://www.repeatmasker.org/>.
126. Smit, A.F.A., and Hubley, R. (2008–2015). RepeatModeler Open-1.0. <https://www.repeatmasker.org/>.
127. Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. *Genome Res.* 14, 1188–1190. <https://doi.org/10.1101/gr.849004>.
128. Cheng, C., Tang, N., Li, J., Cao, S., Zhang, T., Wei, X., and Wang, H. (2019). Bacteria-free minicircle DNA system to generate integration-free CAR-T cells. *J. Med. Genet.* 56, 10–17. <https://doi.org/10.1136/jmedgenet-2018-105405>.
129. Tang, N., Cheng, C., Zhang, X., Qiao, M., Li, N., Mu, W., Wei, X.F., Han, W., and Wang, H. (2020). TGF-beta inhibition via CRISPR promotes the long-term efficacy of CAR T cells against solid tumors. *JCI Insight* 5, e133977. <https://doi.org/10.1172/jci.insight.133977>.
130. Grabundzija, I., Messing, S.A., Thomas, J., Cosby, R.L., Bilic, I., Miskey, C., Gogol-Döring, A., Kapitonov, V., Diem, T., Dalda, A., et al. (2016). A Helitron transposon reconstructed from bats reveals a novel mechanism of genome shuffling in eukaryotes. *Nat. Commun.* 7, 10716. <https://doi.org/10.1038/ncomms10716>.
131. Paatero, A.O., Turakainen, H., Happonen, L.J., Olsson, C., Palomäki, T., Pajunen, M.I., Meng, X., Otonkoski, T., Tuuri, T., Berry, C., et al. (2008). Bacteriophage mu integration in yeast and mammalian genomes. *Nucleic Acids Res.* 36, e148. <https://doi.org/10.1093/nar/gkn801>.
132. Markova, D.N., Ruma, F.B., Casola, C., Mirsalehi, A., and Betrán, E. (2022). Recurrent co-domestication of PIF/Harbinger transposable element proteins in insects. *Mobile DNA* 13, 28. <https://doi.org/10.1186/s13100-022-00282-2>.
133. Katoh, K., and Toh, H. (2008). Recent developments in the MAFFT multiple sequence alignment program. *Brief. Bioinform.* 9, 286–298. <https://doi.org/10.1093/bib/bbn013>.
134. Mitra, R., Li, X., Kapusta, A., Mayhew, D., Mitra, R.D., Feschotte, C., and Craig, N.L. (2013). Functional characterization of piggyBat from the bat *Myotis lucifugus* unveils an active mammalian DNA transposon. *Proc. Natl. Acad. Sci. USA* 110, 234–239. <https://doi.org/10.1073/pnas.1217548110>.
135. Miskey, C., Izsvák, Z., Plasterk, R.H., and Ivics, Z. (2003). The Frog Prince: a reconstructed transposon from *Rana pipiens* with high transpositional activity in vertebrate cells. *Nucleic Acids Res.* 31, 6873–6881. <https://doi.org/10.1093/nar/gkg910>.
136. Mahillon, J., and Chandler, M. (1998). Insertion sequences. *Microbiol. Mol. Biol. Rev.* 62, 725–774. <https://doi.org/10.1128/MMBR.62.3.725-774.1998>.
137. Ho, L., and Ané, C. (2014). A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. *Syst. Biol.* 63, 397–408. <https://doi.org/10.1093/sysbio/syu005>.

138. Ives, A.R., and Garland, T., Jr. (2010). Phylogenetic logistic regression for binary dependent variables. *Syst. Biol.* 59, 9–26. <https://doi.org/10.1093/sysbio/syp074>.
139. Kumar, S., Suleski, M., Craig, J.M., Kasprzowicz, A.E., Sanderford, M., Li, M., Stecher, G., and Hedges, S.B. (2022). TimeTree 5: An expanded resource for species divergence times. *Mol. Biol. Evol.* 39, msac174. <https://doi.org/10.1093/molbev/msac174>.
140. Burnham, K.P., Anderson, D.R., and Burnham, K.P. (2002). *Model Selection and Multimodel Inference: a Practical Information-Theoretic Approach* (Springer).
141. Garland, T., Dickerman, A.W., Janis, C.M., and Jones, J.A. (1993). Phylogenetic analysis of covariance by computer-simulation. *Syst. Biol.* 42, 265–292. <https://doi.org/10.1093/sysbio/42.3.265>.
142. Pennell, M.W., Eastman, J.M., Slater, G.J., Brown, J.W., Uyeda, J.C., FitzJohn, R.G., Alfaro, M.E., and Harmon, L.J. (2014). geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics* 30, 2216–2218. <https://doi.org/10.1093/bioinformatics/btu181>.
143. Spanopoulou, E., Zaitseva, F., Wang, F.H., Santagata, S., Baltimore, D., and Panayotou, G. (1996). The homeodomain region of Rag-1 reveals the parallel mechanisms of bacterial and V(D)J recombination. *Cell* 87, 263–276. [https://doi.org/10.1016/S0092-8674\(00\)81344-6](https://doi.org/10.1016/S0092-8674(00)81344-6).
144. Lohe, A.R., De Aguiar, D., and Hartl, D.L. (1997). Mutations in the mariner transposase: the D,D(35)E consensus sequence is nonfunctional. *Proc. Natl. Acad. Sci. USA* 94, 1293–1297. <https://doi.org/10.1073/pnas.94.4.1293>.
145. Liu, D., and Chalmers, R. (2014). Hyperactive mariner transposons are created by mutations that disrupt allosterism and increase the rate of transposon end synapsis. *Nucleic Acids Res.* 42, 2637–2645. <https://doi.org/10.1093/nar/gkt1218>.
146. Aravind, L. (2000). The BED finger, a novel DNA-binding domain in chromatin-boundary-element-binding proteins and transposases. *Trends Biochem. Sci.* 25, 421–423. [https://doi.org/10.1016/S0968-0004\(00\)01620-0](https://doi.org/10.1016/S0968-0004(00)01620-0).
147. Hickman, A.B., Perez, Z.N., Zhou, L., Musungirimi, P., Ghirlando, R., Hinchshaw, J.E., Craig, N.L., and Dyda, F. (2005). Molecular architecture of a eukaryotic DNA transposase. *Nat. Struct. Mol. Biol.* 12, 715–721. <https://doi.org/10.1038/nsmb970>.
148. Carpenter, R.O., Evbuomwan, M.O., Pittaluga, S., Rose, J.J., Raffeld, M., Yang, S., Gress, R.E., Hakim, F.T., and Kochenderfer, J.N. (2013). B-cell maturation antigen is a promising target for adoptive T-cell therapy of multiple myeloma. *Clin. Cancer Res.* 19, 2048–2060. <https://doi.org/10.1158/1078-0432.CCR-12-2422>.
149. Tsai, S.Q., Zheng, Z., Nguyen, N.T., Liebers, M., Topkar, V.V., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, A.J., Le, L.P., et al. (2015). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* 33, 187–197. <https://doi.org/10.1038/nbt.3117>.
150. Li, H., and Homer, N. (2010). A survey of sequence alignment algorithms for next-generation sequencing. *Brief. Bioinform.* 11, 473–483. <https://doi.org/10.1093/bib/bbq015>.
151. Chakravarty, D., Gao, J., Phillips, S.M., Kundra, R., Zhang, H., Wang, J., Rudolph, J.E., Yaeger, R., Soumerai, T., Nissan, M.H., et al. (2017). OncoKB: A precision oncology knowledge base. *JCO Precis. Oncol.* 2017. <https://doi.org/10.1200/PO.17.00011>.
152. Good, C.R., Aznar, M.A., Kuramitsu, S., Samareh, P., Agarwal, S., Donahue, G., Ishiyama, K., Wellhausen, N., Rennels, A.K., Ma, Y., et al. (2021). An NK-like CAR T cell transition in CAR T cell dysfunction. *Cell* 184, 6081–6100.e26. <https://doi.org/10.1016/j.cell.2021.11.016>.
153. Balciunas, D., Wangenstein, K.J., Wilber, A., Bell, J., Geurts, A., Sivasubbu, S., Wang, X., Hackett, P.B., Largaespada, D.A., McIvor, R.S., and Ekker, S.C. (2006). Harnessing a high cargo-capacity transposon for genetic applications in vertebrates. *PLoS Genet.* 2, e169. <https://doi.org/10.1371/journal.pgen.0020169>.
154. Mátés, L., Chuah, M.K.L., Belay, E., Jerchow, B., Manoj, N., Acosta-Sanchez, A., Grzela, D.P., Schmitt, A., Becker, K., Matrai, J., et al. (2009). Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. *Nat. Genet.* 41, 753–761. <https://doi.org/10.1038/ng.343>.
155. Chee, M.K., and Haase, S.B. (2012). New and redesigned pRS plasmid shuttle vectors for genetic manipulation of *Saccharomyces cerevisiae*. *G3 (Bethesda)* 2, 515–526. <https://doi.org/10.1534/g3.111.001917>.
156. Chicaybam, L., Sodre, A.L., Curzio, B.A., and Bonamino, M.H. (2013). An efficient low cost method for gene transfer to T lymphocytes. *PloS One* 8, e60298. <https://doi.org/10.1371/journal.pone.0060298>.
157. Mestermann, K., Eichler, M., Machwirth, M., Kebbel, K., Köhl, U., Einsele, H., Müller, C., Lehmann, J., Raskó, T., Lundberg, F., et al. (2020). P09. 08 Clinical-grade manufacturing of ROR1 CAR T cells using a novel virus-free protocol. *J. Immunother. Cancer* 8, A55.2–A55.6. <https://doi.org/10.1136/jitc-2020-ITOC7.108>.
158. Gomes-Silva, D., Mukherjee, M., Srinivasan, M., Krenciute, G., Dakhova, O., Zheng, Y., Cabral, J.M.S., Rooney, C.M., Orange, J.S., Brenner, M.K., and Mamontkin, M. (2017). Tonic 4–1BB costimulation in chimeric antigen receptors impedes T cell survival and is vector-dependent. *Cell Rep.* 21, 17–26. <https://doi.org/10.1016/j.celrep.2017.09.015>.
159. Hashimoto, M., Kamphorst, A.O., Im, S.J., Kissick, H.T., Pillai, R.N., Ramalingam, S.S., Araki, K., and Ahmed, R. (2018). CD8 T cell exhaustion in chronic infection and cancer: opportunities for interventions. *Annu. Rev. Med.* 69, 301–318. <https://doi.org/10.1146/annurev-med-012017-043208>.
160. Wherry, E.J. (2011). T cell exhaustion. *Nat. Immunol.* 12, 492–499. <https://doi.org/10.1038/ni.2035>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Pacific Blue™ Mouse Anti-Human CD3	Biolegend	Cat#558117; RRID:AB_397038
PE anti-human CD4 Antibody (clone: OKT4)	Biolegend	Cat#317410; RRID:AB_571954
APC anti-human CD8a Antibody	Biolegend	Cat#301014; RRID:AB_2562054
PE anti-human CD45RO Antibody	Biolegend	Cat#304206; RRID:AB_2564160
APC anti-human CD197 (CCR7) Antibody	Biolegend	Cat#353213; RRID:AB_10915474
Alexa Fluor 647 AffiniPure F(ab') ₂ Fragment Goat Anti-Human IgG (H+L)	Jackson ImmunoResearch	Cat#109-606-003; RRID:AB_2337892
Alexa Fluor 647 AffiniPure Goat Anti-Mouse IgG, F(ab') ₂ Fragment Specific	Jackson ImmunoResearch	Cat#115-605-072; RRID:AB_2338910
PE anti-human CD3 Antibody	Biolegend	Cat#300408; RRID:AB_2564150
Brilliant Violet 421™ anti-human CD279 (PD-1) Antibody	Biolegend	Cat#367422; RRID:AB_2721516
Brilliant Violet 421™ anti-human CD223 (LAG-3) Antibody	Biolegend	Cat#369313; RRID:AB_2629797
APC anti-human CD152 (CTLA-4) Antibody	Biolegend	Cat#349908; RRID:AB_10680785
APC anti-human CD366 (Tim-3) Antibody	Biolegend	Cat#345012; RRID:AB_2561717
APC anti-human CD95 (Fas) Antibody	Biolegend	Cat#305611; RRID:AB_314550
Recombinant Alexa Fluor® 647 Anti-CD3 zeta (phospho Y83) antibody[EP776(2)Y]	abcam	Cat#ab237452; RRID:AB_3099663
Anti-(G4S)n (B02H1) mAb	Hycells	Cat#GS-ARFT100; RRID:AB_3099665
Bacterial and virus strains		
pMD2.G	Addgene	Cat#12259
psPAX2	Addgene	Cat#12260
FUW-EF1α-CAR	This paper	N/A
pRRLSIN-EF1α-CAR	Constructed by our lab	N/A
Biological samples		
Human umbilical cord blood	Beijing Cord Blood Bank	N/A
Chemicals, peptides, and recombinant proteins		
Recombinant human IL2 protein	Sino Biological Inc	Cat#11848-HNAE
Recombinant human IL7 protein	Sino Biological Inc	Cat#11821-HNAE
Recombinant human IL15 protein	Sino Biological Inc	Cat#10360-HNCE
PE-Labeled Human CD19 (20-291) Protein, His Tag (Site-specific conjugation)	ACROBiosystems	Cat#CD9-HP2H3
FITC-Labeled Human EGFR Protein, His Tag DMF Filed	ACROBiosystems	Cat#EGR-HF2H5
7-AAD Viability Staining Solution	Biolegend	Cat#420404
L-Glutamine	Gibco	Cat#25030081
Trypsin-EDTA (0.25%)	Gibco	Cat#25200072
Matrigel® Matrix	Corning	Cat#354277
Penicillin-Streptomycin	Thermo Fisher	Cat#15140-122
Methylene Blue	Solarbio	Cat#M8030
Paraformaldehyde	aladdin	Cat#C104190
Human mononuclear cell separation fluid	DongFang HuaHui Biomedical Technology	Cat#25710
ACK Lysis Buffer	Thermo Fisher	Cat#A1049201

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical commercial assays		
Lipofectamine™ 3000 Transfection Reagent	Thermo Fisher	Cat#L3000015
DNeasy®Blood & Tissue Kit	QIAGEN	Cat#69504
DNA Clean Beads	Vazyme	Cat#N411
DNA Damage Repair Kit	Vazyme	Cat#N208
Universal End preparation Module for illumina	Vazyme	Cat#N203
PrimeSTAR® HS DNA Polymerase	Takara	Cat#R010A
ImmunoSep Human CD3 ⁺ cell positive selection kit	Beijing Nuowei Biotechnology	Cat#710305
Dynabeads™ Human T-Activator CD3/CD28 for T Cell Expansion and Activation	Thermo Fisher	Cat#11131D
P3 Primary Cell 4D-Nucleofector X Kit	Lonza	Cat#V4XP-3024
Steady-Glo® Luciferase Assay System	Promega	Cat#E2520
Deposited data		
20 known active DNA TEs	This paper	Table S1
Sequences of 130 potentially active TEs	This paper	Table S2
TIR classification information	This paper	Table S4
Primers	This paper	Table S5
Raw colony images for transposition activity of 130 TEs in human cells	This paper	Mendeley Data: https://doi.org/10.17632/pms2bvn442.2
Raw sequencing data	This paper	NCBI SRA: PRJNA988388 & NGDC GSA: PRJCA017989
Code for detecting and characterizing DNA TEs, and analyzing insertion sites	This paper	Zenodo: https://doi.org/10.5281/zenodo.8106731
Alignment of <i>Tc1</i> transposases	This paper	Mendeley Data: https://doi.org/10.17632/ktfwtk6k3f.1
Alignment of <i>hAT</i> transposases	This paper	Mendeley Data: https://doi.org/10.17632/b54z6nvy7t.1
Overlaid transposases' structures of SB100X and <i>Tc1-2_ST</i>	This paper	Mendeley Data: https://doi.org/10.17632/bz7kh3cgpw.1
Overlaid Transposases-TIR complexes of <i>Hermes</i> , <i>HAT1_AG</i> and <i>hAT-7_PM</i>	This paper	Mendeley Data: https://doi.org/10.17632/y4xtj2yn7m.1
Scatter plot of non-autonomous TEs relative to the consensus autonomous sequences	This paper	Mendeley Data: https://doi.org/10.17632/bycpd2tcxw.1
The insertion sites of 12 DNA TEs	This paper	Mendeley Data: https://doi.org/10.17632/2bmx6235k7.1
Experimental models: Cell lines		
Human cell line: HEK 293T cell line	ATCC	Cat#CRL-11268
Human cell line: HeLa cell line	ATCC	Cat#CRM-CCL-2
Human cell line: K562 cell line	ATCC	Cat#CRL-3343
Human cell line: NCI-H226 cell line	ATCC	Cat#CRL-5826
Mouse cell line: Neuro-2a	ATCC	Cat#CCL-131
Human cell line: K562-CD19-luciferase cell line	Constructed in our previous paper	PMID: 27910851
Human cell line: Raji-luciferase cell line	Constructed in our previous paper	PMID: 27910851
Human cell line: NCI-H226-luciferase cell line	Constructed in our previous paper	PMID: 31999649

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental models: Organisms/strains		
Mouse: NPG (NOD.Cg-Prkdc ^{scid} Il2rg ^{tm1Vst/Vst})	Vitalstar	N/A
Oligonucleotides		
Recombinant DNA		
Plasmid: pCMV-hAT-2_AG	This study	Table S3
Plasmid: pCMV-AgaP12	This study	Table S3
Plasmid: pCMV-P3_AG	This study	Table S3
Plasmid: pCMV-HAT2_CI	This study	Table S3
Plasmid: pCMV-HAT5_CI	This study	Table S3
Plasmid: pCMV-hAT-6_DR	This study	Table S3
Plasmid: pCMV-Chaplin1_DR	This study	Table S3
Plasmid: pCMV-Harbinger-4_XT-Myb	This study	Table S3
Plasmid: pCMV-Harbinger-4_XT-Tn	This study	Table S3
Plasmid: pCMV-HAT1_AG	This study	Table S3
Plasmid: pCMV-AgaP15	This study	Table S3
Plasmid: pCMV-IS4EU-1_DR	This study	Table S3
Plasmid: pCMV-POGO	This study	Table S3
Plasmid: pCMV-Tc1-8B_DR	This study	Table S3
Plasmid: pCMV-HOBO	This study	Table S3
Plasmid: pCMV-hAT-6_PM	This study	Table S3
Plasmid: pCMV-MARISP1	This study	Table S3
Plasmid: pCMV-BARI_DM	This study	Table S3
Plasmid: pCMV-hAT-3_XT	This study	Table S3
Plasmid: pCMV-P1_AG	This study	Table S3
Plasmid: pCMV-IS4EU-2_DR-Myb	This study	Table S3
Plasmid: pCMV-IS4EU-2_DR-Tn	This study	Table S3
Plasmid: pCMV-Tc1-3_Xt	This study	Table S3
Plasmid: pCMV-Mariner-1_PM	This study	Table S3
Plasmid: pCMV-P2_AG	This study	Table S3
Plasmid: pCMV-hAT-5_DR	This study	Table S3
Plasmid: pCMV-Tc1-3_FR	This study	Table S3
Plasmid: pCMV-Mariner-4_XT	This study	Table S3
Plasmid: pCMV-Tc1-5_Xt	This study	Table S3
Plasmid: pCMV-Tc1-10_Xt	This study	Table S3
Plasmid: pCMV-hAT-1_DP	This study	Table S3
Plasmid: pCMV-Mariner2_AG	This study	Table S3
Plasmid: pCMV-Tc1-1_Xt	This study	Table S3
Plasmid: pCMV-Tc1-1_AG	This study	Table S3
Plasmid: pCMV-Mariner-1_XT	This study	Table S3
Plasmid: pCMV-Tc1DR3_Xt	This study	Table S3
Plasmid: pCMV-hAT-1B_PM	This study	Table S3
Plasmid: pCMV-hAT-3_PM	This study	Table S3
Plasmid: pCMV-hAT-6B_PM	This study	Table S3
Plasmid: pCMV-Tc1-1_PM	This study	Table S3
Plasmid: pCMV-TC1_XL	This study	Table S3
Plasmid: pCMV-Mariner-4_AMi	This study	Table S3
Plasmid: pCMV-Myotis_hAT1	This study	Table S3

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Plasmid: pCMV-hAT-7_PM	This study	Table S3
Plasmid: pCMV-TC1_FR4	This study	Table S3
Plasmid: pCMV-hAT-8_PM	This study	Table S3
Plasmid: pCMV-piggyBac1_Mm	This study	Table S3
Plasmid: pCMV-Tc1-11_Xt	This study	Table S3
Plasmid: pCMV-Tc1-16_Xt	This study	Table S3
Plasmid: pCMV-TC1-2_DM	This study	Table S3
Plasmid: pCMV-Tc1-4_Xt	This study	Table S3
Plasmid: pCMV-TC1_DM	This study	Table S3
Plasmid: pCMV-Tc1-15_Xt	This study	Table S3
Plasmid: pCMV-TC1_FR2	This study	Table S3
Plasmid: pCMV-Mariner-6B_AMi	This study	Table S3
Plasmid: pCMV-Tc1-9_Xt	This study	Table S3
Plasmid: pCMV-hAT-9_XT	This study	Table S3
Plasmid: pCMV-Mariner-5_XT	This study	Table S3
Plasmid: pCMV-S2_DM	This study	Table S3
Plasmid: pCMV-PROTOP	This study	Table S3
Plasmid: pCMV-Tc1-8_Xt	This study	Table S3
Plasmid: pCMV-Tc1-12_Xt	This study	Table S3
Plasmid: pCMV-Mariner-6_AMi	This study	Table S3
Plasmid: pCMV-hAT-3_Gav	This study	Table S3
Plasmid: pCMV-Mariner-2_XT	This study	Table S3
Plasmid: pCMV-hAT-4_Crp	This study	Table S3
Plasmid: pCMV-OposCharlie2	This study	Table S3
Plasmid: pCMV-hAT-1_AMi	This study	Table S3
Plasmid: pCMV-Mariner-2_AMi	This study	Table S3
Plasmid: pCMV-hAT-13_AMi	This study	Table S3
Plasmid: pCMV-hAT-12_AMi	This study	Table S3
Plasmid: pCMV-Mariner-7_Croc	This study	Table S3
Plasmid: pCMV-piggyBac-2_XT	This study	Table S3
Plasmid: pCMV-Mariner-3_AMi	This study	Table S3
Plasmid: pCMV-piggyBac1_CI	This study	Table S3
Plasmid: pCMV-piggyBac-1_AMi	This study	Table S3
Plasmid: pCMV-Mariner-5_AMi	This study	Table S3
Plasmid: pCMV-Mariner-6_Crp	This study	Table S3
Plasmid: pCMV-Mariner-3_Crp	This study	Table S3
Plasmid: pCMV-hAT-3_AMi	This study	Table S3
Plasmid: pCMV-TC1_FR1	This study	Table S3
Plasmid: pCMV-hAT-5_Croc	This study	Table S3
Plasmid: pCMV-hAT-8_AMi	This study	Table S3
Plasmid: pCMV-PARIS	This study	Table S3
Plasmid: pCMV-Mariner-2_PM	This study	Table S3
Plasmid: pCMV-piggyBac-1_XT	This study	Table S3
Plasmid: pCMV-Tigger1	This study	Table S3
Plasmid: pCMV-Mariner-1_Crp	This study	Table S3
Plasmid: pCMV-S_DM	This study	Table S3
Plasmid: pCMV-hAT-12_Crp	This study	Table S3
Plasmid: pCMV-hAT-1_PM	This study	Table S3

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Plasmid: pCMV-Senkusha1	This study	Table S3
Plasmid: pCMV-Tc1-2_PM	This study	Table S3
Plasmid: pCMV-hAT-2_Gav	This study	Table S3
Plasmid: pCMV-Tigger2f	This study	Table S3
Plasmid: pCMV-Tc1-4_DR	This study	Table S3
Plasmid: pCMV-Tigger17	This study	Table S3
Plasmid: pCMV-Tigger3	This study	Table S3
Plasmid: pCMV-hAT-14_Crp	This study	Table S3
Plasmid: pCMV-Mariner-9_Crp	This study	Table S3
Plasmid: pCMV-hAT-11_AMi	This study	Table S3
Plasmid: pCMV-hAT-17_Croc	This study	Table S3
Plasmid: pCMV-hAT-4_AMi	This study	Table S3
Plasmid: pCMV-Tigger4	This study	Table S3
Plasmid: pCMV-Tigger7	This study	Table S3
Plasmid: pCMV-Tigger2	This study	Table S3
Plasmid: pCMV-hAT-19_Crp	This study	Table S3
Plasmid: pCMV-hAT-17B_Croc	This study	Table S3
Plasmid: pCMV-Tc1-13_Xt	This study	Table S3
Plasmid: pCMV-hAT-19B_Croc	This study	Table S3
Plasmid: pCMV-MarsTigger8	This study	Table S3
Plasmid: pCMV-Tigger5	This study	Table S3
Plasmid: pCMV-hAT-10_XT	This study	Table S3
Plasmid: pCMV-hAT-6_AMi	This study	Table S3
Plasmid: pCMV-MarsTigger1c	This study	Table S3
Plasmid: pCMV-Tigger17c	This study	Table S3
Plasmid: pCMV-Tc1-14_Xt	This study	Table S3
Plasmid: pCMV-Kanga1	This study	Table S3
Plasmid: pCMV-Arthur1	This study	Table S3
Plasmid: pCMV-Harbinger-1_Crp-Myb	This study	Table S3
Plasmid: pCMV-Harbinger-1_Crp-Tn	This study	Table S3
Plasmid: pCMV-Zaphod3	This study	Table S3
Plasmid: pCMV-Joey1	This study	Table S3
Plasmid: pCMV-Zaphod	This study	Table S3
Plasmid: pCMV-Harbinger-3_Ami-Myb	This study	Table S3
Plasmid: pCMV-Harbinger-3_Ami-Tn	This study	Table S3
Plasmid: pCMV-MARWOLEN1	This study	Table S3
Plasmid: pCMV-Zaphod2	This study	Table S3
Plasmid: pCMV-Tc1-1_PS	This study	Table S3
Plasmid: pCMV-Tc1-1_ST	This study	Table S3
Plasmid: pCMV-Tc1-3_ST	This study	Table S3
Plasmid: pCMV-Tc1-4_ST	This study	Table S3
Plasmid: pCMV-Tc1-5_ST	This study	Table S3
Plasmid: pCMV-Tc1-2_ST	This study	Table S3
Plasmid: pCMV-piggyBac1_Mm	This study	Table S3
Plasmid: pCMV-piggyBac-2_XT	This study	Table S3
Plasmid: pCMV-piggyBac1_CI	This study	Table S3
Plasmid: pCMV-piggyBac-1_AMi	This study	Table S3
Plasmid: pCMV-piggyBac-1_XT	This study	Table S3

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Plasmid: pMV-hAT-2_AG	This study	Table S3
Plasmid: pMV-AgaP12	This study	Table S3
Plasmid: pMV-P3_AG	This study	Table S3
Plasmid: pMV-HAT2_CI	This study	Table S3
Plasmid: pMV-HAT5_CI	This study	Table S3
Plasmid: pMV-hAT-6_DR	This study	Table S3
Plasmid: pMV-Chaplin1_DR	This study	Table S3
Plasmid: pMV-Harbinger-4_XT	This study	Table S3
Plasmid: pMV-HAT1_AG	This study	Table S3
Plasmid: pMV-AgaP15	This study	Table S3
Plasmid: pMV-IS4EU-1_DR	This study	Table S3
Plasmid: pMV-POGO	This study	Table S3
Plasmid: pMV-Tc1-8B_DR	This study	Table S3
Plasmid: pMV-HOBO	This study	Table S3
Plasmid: pMV-hAT-6_PM	This study	Table S3
Plasmid: pMV-MARISP1	This study	Table S3
Plasmid: pMV-BARI_DM	This study	Table S3
Plasmid: pMV-hAT-3_XT	This study	Table S3
Plasmid: pMV-P1_AG	This study	Table S3
Plasmid: pMV-IS4EU-2_DR	This study	Table S3
Plasmid: pMV-Tc1-3_Xt	This study	Table S3
Plasmid: pMV-Mariner-1_PM	This study	Table S3
Plasmid: pMV-P2_AG	This study	Table S3
Plasmid: pMV-hAT-5_DR	This study	Table S3
Plasmid: pMV-Tc1-3_FR	This study	Table S3
Plasmid: pMV-Mariner-4_XT	This study	Table S3
Plasmid: pMV-Tc1-5_Xt	This study	Table S3
Plasmid: pMV-Tc1-10_Xt	This study	Table S3
Plasmid: pMV-hAT-1_DP	This study	Table S3
Plasmid: pQLL-Mariner2_AG	This study	Table S3
Plasmid: pMV-Tc1-1_Xt	This study	Table S3
Plasmid: pQLL-Tc1-1_AG	This study	Table S3
Plasmid: pMV-Mariner-1_XT	This study	Table S3
Plasmid: pMV-Tc1DR3_Xt	This study	Table S3
Plasmid: pMV-hAT-1B_PM	This study	Table S3
Plasmid: pMV-hAT-3_PM	This study	Table S3
Plasmid: pMV-hAT-6B_PM	This study	Table S3
Plasmid: pMV-Tc1-1_PM	This study	Table S3
Plasmid: pMV-TC1_XL	This study	Table S3
Plasmid: pMV-Mariner-4_AMi	This study	Table S3
Plasmid: pMV-Myotis_hAT1	This study	Table S3
Plasmid: pMV-hAT-7_PM	This study	Table S3
Plasmid: pMV-TC1_FR4	This study	Table S3
Plasmid: pMV-hAT-8_PM	This study	Table S3
Plasmid: pMV-piggyBac1_Mm	This study	Table S3
Plasmid: pMV-Tc1-11_Xt	This study	Table S3
Plasmid: pMV-Tc1-16_Xt	This study	Table S3
Plasmid: pMV-TC1-2_DM	This study	Table S3

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Plasmid: pMV-Tc1-4_Xt	This study	Table S3
Plasmid: pMV-TC1_DM	This study	Table S3
Plasmid: pMV-Tc1-15_Xt	This study	Table S3
Plasmid: pMV-TC1_FR2	This study	Table S3
Plasmid: pMV-Mariner-6B_AMi	This study	Table S3
Plasmid: pMV-Tc1-9_Xt	This study	Table S3
Plasmid: pMV-hAT-9_XT	This study	Table S3
Plasmid: pMV-Mariner-5_XT	This study	Table S3
Plasmid: pMV-S2_DM	This study	Table S3
Plasmid: pMV-PROTOP	This study	Table S3
Plasmid: pMV-Tc1-8_Xt	This study	Table S3
Plasmid: pMV-Tc1-12_Xt	This study	Table S3
Plasmid: pMV-Mariner-6_AMi	This study	Table S3
Plasmid: pMV-hAT-3_Gav	This study	Table S3
Plasmid: pMV-Mariner-2_XT	This study	Table S3
Plasmid: pMV-hAT-4_Crp	This study	Table S3
Plasmid: pMV-OposCharlie2	This study	Table S3
Plasmid: pMV-hAT-1_AMi	This study	Table S3
Plasmid: pMV-Mariner-2_AMi	This study	Table S3
Plasmid: pMV-hAT-13_AMi	This study	Table S3
Plasmid: pMV-hAT-12_AMi	This study	Table S3
Plasmid: pMV-Mariner-7_Croc	This study	Table S3
Plasmid: pMV-piggyBac-2_XT	This study	Table S3
Plasmid: pMV-Mariner-3_AMi	This study	Table S3
Plasmid: pMV-piggyBac1_CI	This study	Table S3
Plasmid: pMV-piggyBac-1_AMi	This study	Table S3
Plasmid: pMV-Mariner-5_AMi	This study	Table S3
Plasmid: pMV-Mariner-6_Crp	This study	Table S3
Plasmid: pMV-Mariner-3_Crp	This study	Table S3
Plasmid: pMV-hAT-3_AMi	This study	Table S3
Plasmid: pMV-TC1_FR1	This study	Table S3
Plasmid: pMV-hAT-5_Croc	This study	Table S3
Plasmid: pMV-hAT-8_AMi	This study	Table S3
Plasmid: pMV-PARIS	This study	Table S3
Plasmid: pMV-Mariner-2_PM	This study	Table S3
Plasmid: pQLL-piggyBac-1_XT	This study	Table S3
Plasmid: pMV-Tigger1	This study	Table S3
Plasmid: pMV-Mariner-1_Crp	This study	Table S3
Plasmid: pMV-S_DM	This study	Table S3
Plasmid: pMV-hAT-12_Crp	This study	Table S3
Plasmid: pMV-hAT-1_PM	This study	Table S3
Plasmid: pMV-Senkusha1	This study	Table S3
Plasmid: pMV-Tc1-2_PM	This study	Table S3
Plasmid: pMV-hAT-2_Gav	This study	Table S3
Plasmid: pMV-Tigger2f	This study	Table S3
Plasmid: pMV-Tc1-4_DR	This study	Table S3
Plasmid: pMV-Tigger17	This study	Table S3
Plasmid: pMV-Tigger3	This study	Table S3

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Plasmid: pMV-hAT-14_Crp	This study	Table S3
Plasmid: pMV-Mariner-9_Crp	This study	Table S3
Plasmid: pMV-hAT-11_AMi	This study	Table S3
Plasmid: pMV-hAT-17_Croc	This study	Table S3
Plasmid: pMV-hAT-4_AMi	This study	Table S3
Plasmid: pMV-Tigger4	This study	Table S3
Plasmid: pMV-Tigger7	This study	Table S3
Plasmid: pMV-Tigger2	This study	Table S3
Plasmid: pMV-hAT-19_Crp	This study	Table S3
Plasmid: pMV-hAT-17B_Croc	This study	Table S3
Plasmid: pMV-Tc1-13_Xt	This study	Table S3
Plasmid: pMV-hAT-19B_Croc	This study	Table S3
Plasmid: pMV-MarsTigger8	This study	Table S3
Plasmid: pMV-Tigger5	This study	Table S3
Plasmid: pQLL-hAT-10_XT	This study	Table S3
Plasmid: pMV-hAT-6_AMi	This study	Table S3
Plasmid: pMV-MarsTigger1c	This study	Table S3
Plasmid: pMV-Tigger17c	This study	Table S3
Plasmid: pMV-Tc1-14_Xt	This study	Table S3
Plasmid: pMV-Kanga1	This study	Table S3
Plasmid: pMV-Arthur1	This study	Table S3
Plasmid: pMV-Harbinger-1_Crp	This study	Table S3
Plasmid: pMV-Zaphod3	This study	Table S3
Plasmid: pMV-Joey1	This study	Table S3
Plasmid: pMV-Zaphod	This study	Table S3
Plasmid: pMV-Harbinger-3_AMi	This study	Table S3
Plasmid: pMV-MARWOLEN1	This study	Table S3
Plasmid: pMV-Zaphod2	This study	Table S3
Plasmid: pMV-Tc1-1_PS	This study	Table S3
Plasmid: pMV-Tc1-1_ST	This study	Table S3
Plasmid: pMV-Tc1-2_ST	This study	Table S3
Plasmid: pQLL-Tc1-3_ST	This study	Table S3
Plasmid: pMV-Tc1-4_ST	This study	Table S3
Plasmid: pMV-Tc1-5_ST	This study	Table S3
Plasmid: pQLL-MAG w/o spacer	This study	Table S3
Plasmid: pQLL-Tc1-1_Xt w/o spacer	This study	Table S3
Plasmid: pQLL-Tc1-3_FR w/o spacer	This study	Table S3
Plasmid: pQLL-Tc1-8B_DR w/o spacer	This study	Table S3
Plasmid: pMV-Tc1-1_PM w/o spacer	This study	Table S3
Plasmid: pMV-Tc1-2_ST w/o spacer	This study	Table S3
Plasmid: pMV-Tc1-2_ST DRi ^{KO}	This study	Table S3
Plasmid: pCMV-SB100X-T102A	This study	Table S3
Plasmid: pCMV-SB100X-R131Y	This study	Table S3
Plasmid: pCMV-SB100X-E154A	This study	Table S3
Plasmid: pCMV-SB100X-W195M	This study	Table S3
Plasmid: pCMV-SB100X-Y218F	This study	Table S3
Plasmid: pCMV-SB100X-H249A	This study	Table S3
Plasmid: pCMV-SB100X-W268A	This study	Table S3

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Plasmid: pCMV-SB100X-P269A	This study	Table S3
Plasmid: pCMV-SB100X-P277A	This study	Table S3
Plasmid: pCMV-SB100X-I278A	This study	Table S3
Plasmid: pCMV-SB100X-W282F	This study	Table S3
Plasmid: pCMV-SB100X-W308A	This study	Table S3
Plasmid: pCMV-SB100X-L319A	This study	Table S3
Plasmid: pCMV-SB100X-R326H	This study	Table S3
Plasmid: pCMV-SB100X-G335A	This study	Table S3
Plasmid: pCMV-Tc1-2_ST-T102A	This study	Table S3
Plasmid: pCMV-Tc1-2_ST-R131Y	This study	Table S3
Plasmid: pCMV-Tc1-2_ST-E154A	This study	Table S3
Plasmid: pCMV-Tc1-2_ST-W195M	This study	Table S3
Plasmid: pCMV-Tc1-2_ST-Y218F	This study	Table S3
Plasmid: pCMV-Tc1-2_ST-H249A	This study	Table S3
Plasmid: pCMV-Tc1-2_ST-W268A	This study	Table S3
Plasmid: pCMV-Tc1-2_ST-P269A	This study	Table S3
Plasmid: pCMV-Tc1-2_ST-P277A	This study	Table S3
Plasmid: pCMV-Tc1-2_ST-I278A	This study	Table S3
Plasmid: pCMV-Tc1-2_ST-W282F	This study	Table S3
Plasmid: pCMV-Tc1-2_ST-W308A	This study	Table S3
Plasmid: pCMV-Tc1-2_ST-L319A	This study	Table S3
Plasmid: pCMV-Tc1-2_ST-R326H	This study	Table S3
Plasmid: pCMV-Tc1-2_ST-G335A	This study	Table S3
Plasmid: pCMV-HAT1_AG-C265A	This study	Table S3
Plasmid: pCMV-HAT1_AG-H268A	This study	Table S3
Plasmid: pCMV-HAT1_AG-R318A	This study	Table S3
Plasmid: pCMV-HAT1_AG-S321A	This study	Table S3
Plasmid: pCMV-HAT1_AG-P437A	This study	Table S3
Plasmid: pCMV-HAT1_AG-S576A	This study	Table S3
Plasmid: pCMV-HAT1_AG-R586A	This study	Table S3
Plasmid: pCMV-hAT-7-PM-C265A	This study	Table S3
Plasmid: pCMV-hAT-7-PM-H268A	This study	Table S3
Plasmid: pCMV-hAT-7-PM-R318A	This study	Table S3
Plasmid: pCMV-hAT-7-PM-S321A	This study	Table S3
Plasmid: pCMV-hAT-7-PM-P437A	This study	Table S3
Plasmid: pCMV-hAT-7-PM-S576A	This study	Table S3
Plasmid: pCMV-hAT-7-PM-R586A	This study	Table S3
Plasmid: pICoZ-MAG-K16V	This study	Table S3
Plasmid: pICoZ-MAG-E30K	This study	Table S3
Plasmid: pICoZ-MAG-N102G	This study	Table S3
Plasmid: pICoZ-MAG-N152E	This study	Table S3
Plasmid: pICoZ-MAG-V154F	This study	Table S3
Plasmid: pICoZ-MAG-R163W	This study	Table S3
Plasmid: pICoZ-MAG-P165K	This study	Table S3
Plasmid: pICoZ-MAG-A190G	This study	Table S3
Plasmid: pICoZ-MAG-I200L	This study	Table S3
Plasmid: pICoZ-MAG-M235Q	This study	Table S3
Plasmid: pICoZ-MAG-A263S	This study	Table S3

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Plasmid: pICoZ-MAG-H181A	This study	Table S3
Plasmid: pICoZ-MAG-H181V	This study	Table S3
Plasmid: pICoZ-MAG-P240A	This study	Table S3
Plasmid: pICoZ-MAG-P240R	This study	Table S3
Plasmid: pICoZ-MAG-K241R	This study	Table S3
Plasmid: pMV-SB100X-5 kb	This study	Table S3
Plasmid: pMV-Tc1-2_ST-5 kb	This study	Table S3
Plasmid: pQLL-MAG-5 kb	This study	Table S3
Plasmid: pQLL-Tc1-1_Xt-5 kb	This study	Table S3
Plasmid: pQLL-Tc1-3_FR-5 kb	This study	Table S3
Plasmid: pQLL-Tc1-8B_DR-5 kb	This study	Table S3
Plasmid: pMV-Tc1-1_PM-5 kb	This study	Table S3
Plasmid: pMV-piggyBac-1_AMi-5 kb	This study	Table S3
Plasmid: pMV-HAT1_AG-5 kb	This study	Table S3
Plasmid: pMV-hAT-7_PM-5 kb	This study	Table S3
Plasmid: pMV-SB100X-10 kb	This study	Table S3
Plasmid: pMV-Tc1-2_ST-10 kb	This study	Table S3
Plasmid: pQLL-MAG-10 kb	This study	Table S3
Plasmid: pQLL-Tc1-1_Xt-10 kb	This study	Table S3
Plasmid: pQLL-Tc1-3_FR-10 kb	This study	Table S3
Plasmid: pQLL-Tc1-8B_DR-10 kb	This study	Table S3
Plasmid: pMV-Tc1-1_PM-10 kb	This study	Table S3
Plasmid: pMV-piggyBac-1_AMi-10 kb	This study	Table S3
Plasmid: pMV-HAT1_AG-10 kb	This study	Table S3
Plasmid: pMV-hAT-7_PM-10 kb	This study	Table S3
Plasmid: pRS313	This study	Table S3
Plasmid: pMV-SB100X[EF1 α -copGFP]	This study	Table S3
Plasmid: pQLL-MAG[EF1 α -copGFP]	This study	Table S3
Plasmid: pQLL-Tc1-1_Xt[EF1 α -copGFP]	This study	Table S3
Plasmid: pQLL-Tc1-3_FR[EF1 α -copGFP]	This study	Table S3
Plasmid: pQLL-Tc1-8B_DR[EF1 α -copGFP]	This study	Table S3
Plasmid: pICoZ-SB100X [EF1 α -CD19 4-1BB]	This study	Table S3
Plasmid: pICoZ-MAG[EF1 α -CD19 4-1BB]	This study	Table S3
Plasmid: pICoZ-SB100X [EF1 α -HER1 4-1BB]	This study	Table S3
Plasmid: pICoZ-MAG[EF1 α -HER1 4-1BB]	This study	Table S3
Plasmid: pICoZ-SB100X [EF1 α -CD19 CD28Z]	This study	Table S3
Plasmid: pICoZ-MAG[EF1 α -CD19 CD28Z]	This study	Table S3
Plasmid: pICoZ-SB100X [EF1 α -HER1 CD28Z]	This study	Table S3
Plasmid: pICoZ-MAG[EF1 α -HER1 CD28Z]	This study	Table S3
Plasmid: FUW-[EF1 α -CD19 4-1BB]	This study	Table S3
Plasmid: FUW-[EF1 α -HER1 4-1BB]	This study	Table S3
Plasmid: FUW-[EF1 α -CD19 CD28Z]	This study	Table S3
Plasmid: FUW-[EF1 α -HER1 CD28Z]	This study	Table S3
Plasmid: pRRLSIN-[EF1 α -CD19 4-1BB]	This study	Table S3

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Plasmid: pRRLSIN-[EF1 α -HER1 4-1BB]	This study	Table S3
Plasmid: FUW-[EF1 α -CD19 4-1BB-eGFP]	This study	Table S3
Plasmid: pICOZ-SB100X [EF1 α -CD19 4-1BB-eGFP]	This study	Table S3
Plasmid: pICOZ-MAG [EF1 α -CD19 4-1BB-eGFP]	This study	Table S3
Plasmid: pMD2.G	Addgene	Addgene: 12259
Plasmid: psPAX2	Addgene	Addgene: 12260

Software and algorithms

AlphaFold v2.0.0	Jumper et al. ¹¹²	https://www.deepmind.com/open-source/alphafold
ape v5.5	Paradis et al. ¹¹³	http://ape-package.ird.fr/
bedtools v2.30.0	Quinlan and Hall ¹¹⁴	https://github.com/arq5x/bedtools2
Biorender	Biorender	https://www.biorender.com
ChimeraX v1.4	Pettersen et al. ¹¹⁵	https://www.cgl.ucsf.edu/chimerax/
Cutadapt v3.4	Martin ¹¹⁶	https://cutadapt.readthedocs.io/
ENCODE ATAC-seq pipeline v2.2.2	Hitz et al. ¹¹⁷	https://github.com/ENCODE-DCC/atac-seq-pipeline
ESM1b	Brandes et al. ¹⁰³	https://github.com/ntranoslab/esm-variants/
FlowJo V10	FlowJo LLC	https://www.flowjo.com/solutions/flowjo
GENESCAN v1.0	Burge and Karlin ¹¹⁸	http://argonaute.mit.edu/GENSCAN.html
GraphPad Prism v8	GraphPad Software Inc	https://www.graphpad.com/
ImageJ V1.48	Image Processing and Analysis in Java	https://imagej.net/ij/index.html
Image Lab	Bio-rad	https://www.bio-rad.com/ja-jp/product/image-lab-software
IQ-TREE v2.1.4-beta	Minh et al. ¹¹⁹	http://www.iqtree.org/
iTOL v6	Letunic and Bork ¹²⁰	https://itol.embl.de/
MAFFT v7.407	Yamada et al. ¹²¹	https://mafft.cbrc.jp/alignment/software/
MEME Suite v5.3.0	Bailey et al. ¹²²	https://meme-suite.org/meme/
Novoalign v3.09.04	Novocraft Technologies	http://novocraft.com/
ORFfinder v0.4.3	NCBI	https://ftp.ncbi.nlm.nih.gov/genomes/TOOLS/ORFfinder/
PfamScan v1.6	Madeira et al. ¹²³	http://ftp.ebi.ac.uk/pub/databases/Pfam/Tools/
Picard v2.26.1	Broad Institute	https://broadinstitute.github.io/picard/
pIRS v2.0.2	Hu et al. ¹²⁴	https://github.com/galaxy001/pirs
PolyPhen-2 v2.2.3	Adzhubei et al. ¹⁰²	http://genetics.bwh.harvard.edu/pph2/
PyMOL v 2.5.2	Schrödinger, LLC.	https://pymol.org/
R v4.1.0	R Foundation for Statistical Computing	https://www.R-project.org/
RepeatMasker v4.1.2	Smit et al. ¹²⁵	https://www.repeatmasker.org/
RepeatModeler v1.0.11	Smit and Hubley ¹²⁶	https://www.repeatmasker.org/
WebLogo v3.7.8	Crooks et al. ¹²⁷	http://weblogo.threeplusone.com/

Others

Fetal Bovine Serum	Gibco	Cat#A3161001
DMEM	Gibco	Cat#C11995500CP
RPMI 1640	Gibco	Cat#C11875500CP
Opti-MEM	Gibco	Cat#31985070
DPBS	Gibco	Cat#C14190500CP
CTS OpTmizer™ T Cell Expansion SFM	Gibco	Cat# A1048501

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Yong E. Zhang (zhangyong@ioz.ac.cn).

Materials availability

Almost all plasmids have been uploaded to Addgene for public accessibility as of the publication date.

Data and code availability

- Raw sequencing data have been deposited at the NCBI Sequence Read Archive and the National Genomics Data Center (part of the China National Center for Bioinformation) and are publicly accessible. Original colony images, transposase sequence alignment, predicted transposase structures, three models of evolutionary dynamics and insertion sites of 12 TEs have been deposited at Mendeley and are publicly available. Accession numbers or DOIs are listed in the [key resources table](#).
- All original code has been deposited at Zenodo and is publicly available. Accession numbers or DOIs for these resources are listed in the [key resources table](#).
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cell lines

Depending on the specific purpose, we utilized a total of eight cell lines. Five cell lines, namely HEK293T, HeLa, K562, H266 and Neuro-2a cells, were obtained from the American Type Culture Collection (ATCC). Additionally, K562-CD19-luciferase cells, Raji-luciferase cells (Raji-luci), and H266-luciferase (H266-luci) cells were generated through lentiviral transduction in our previous work.^{128,129} All cell lines were maintained under standard conditions recommended by ATCC. HEK293T, HeLa and Neuro-2a cells were cultured in regular DMEM medium supplemented with 10% (v/v) FBS and 100 µg/mL penicillin and streptomycin. K562 and H266 cells were cultured in RPMI 1640 medium supplemented with 10% FBS and 100 µg/mL penicillin and streptomycin. All cell lines were grown in a humidified 37°C incubator with 5% CO₂. To avoid mycoplasma contamination, monthly testing using the MycoBlue Mycoplasma Detector (Vazyme) was performed on all cell lines.

Primary human T cells

The primary human T cells utilized in this study were isolated from fresh umbilical cord blood, obtained from healthy volunteer donors who provided informed consent; and the cord blood was acquired from the Beijing Cord Blood Bank (Beijing, China). Mononuclear cells were collected from the blood using Histopaque-1077 (Sigma-Aldrich) gradient separation following the manufacturer's instructions. Subsequently, T cells were isolated with the EasySep Human T Cell Enrichment Kit (Stemcell Technologies) and activated with anti-CD3/CD28 Dynabeads (Thermo Fisher Scientific) at a 1:1 ratio. The activated T cells were cultured in OpTmizer™ CTS™ medium (Gibco, ThermoFisher) supplemented with 10% (v/v) heat-inactivated FBS (Gibco), 100 IU/mL IL-2, 10 ng/mL IL-7, and 50 ng/mL IL-15 (SinoBiological). The T cells were maintained in a humidified 37°C incubator with 5% CO₂.

Animals

Tumor xenograft studies were performed using 5- to 6-week-old immunodeficient NOD-Prkdcscid Il2rgtm1Vst/Vst (NPG) female mice (Beijing Vitastar Biotechnology Co., Ltd.). All mice were maintained under pathogen-free conditions in an animal facility (Institute of Zoology, Chinese Academy of Sciences) and cared for in accordance with the policies and certifications of the International Association for Assessment and Accreditation of Laboratory Animal Care.

METHOD DETAILS

Literature survey of DNA TEs active in mammalian cells

We identified 22 transpositionally active DNA TEs in mammalian cells through a literature review. In brief, we searched PubMed with the following keywords: "DNA transposons", "genetic tool" and "mammalian cell". We then manually curated approximately 180 retrieved studies.

Among 22 TEs, 20 eukaryotic TEs belonged to the DNA TEs associated with TIRs. Detailed information regarding these 20 DNA TEs, including their respective superfamily, host species, and references reporting their transpositional activity in mammalian cells, is provided in [Table S1](#). We excluded two TEs including *Helraiser*, harbored by bat genomes and belonging to the *Helitron* group (lacking TIRs), and *Mu*, derived from bacteriophages.^{130,131}

Identification of potentially active DNA TEs

Identification of 124 active TEs from 100 annotated animal genomes

First, we identified potentially full-length autonomous DNA TE copies. We downloaded the genome sequences of 100 animal genomes with TE annotations provided by the UCSC Genome Browser (<http://genome.ucsc.edu/>, as October 2018). By exploiting TE annotation generated via RepeatMasker v4.1.2¹²⁵ searches against the Repbase library (version 24.02), we identified 26,853,019 copies of 2024 DNA TEs across these animal genomes. To identify potentially full-length autonomous TE copies, we focused on those matching over 90% of the consensus sequences provided by Repbase and mapping to the last bases at both ends of consensus. This resulted in a set of 1,895,466 potentially full-length autonomous TE copies, corresponding to 1,577 consensus DNA TEs from Repbase. We purchased the license of Repbase to retrieve the corresponding consensus sequences for further analysis.

Since Repbase consensus sequences lack coding sequence information, we performed annotation. For TEs belonging to the *Tc1/mariner*, *hAT*, and *piggyBac* superfamilies encoding single-exon genes,⁵⁵ we utilized ORFfinder v0.4.3 (<https://www.ncbi.nlm.nih.gov/orffinder/>) to predict open reading frames (ORFs) in their consensus sequences, setting a conservative length cutoff of 300 amino acids (aa). For TEs from the *P* and *PIF/Harbinger* superfamilies, which encode genes with exon–intron structures, we employed GENESCAN v1.0¹¹⁸ to identify ORFs using default parameters. We confirmed the canonical start codons (ATG) and splicing sites (GT..AG) within the ORFs. To identify functional domains, we searched the predicted proteins against the Pfam library using PfamScan v1.6.¹²³ The predicted domains of all proteins included a transposase domain (PF03221, PF05699, PF12017, PF13359, PF13843, and PF01498), along with a myb domain (PF01498, PF05485, and PF05699) specific to the *PIF/Harbinger* superfamily with dual ORFs.^{33,132}

Similarly, we annotated target site duplications (TSDs) and terminal inverted repeats (TIRs). Notably, for each candidate TE region, the UCSC database only annotated one possible TE. However, this region could be derived from another similar TE. To increase the copy number and thus to collect more information for inferring TSDs or TIRs, we performed reannotation by searching each of the 1,577 consensus sequences from Repbase against the 100 genomes with RepeatMasker. TE copies exhibiting identical termini to the consensus TEs were retrieved, and a multiple sequence alignment of the 20-bp sequences flanking each copy was built with MAFFT v7.407¹²¹ given its accuracy.¹³³ TSDs were inferred on the basis of alignment with in-house Perl scripts. For 14 TEs (*Harbinger-1_Crp*, *Harbinger-3_AMi*, *Harbinger-4_XT*, *Senkusha1*, *IS4EU-1_DR*, *IS4EU-2_DR*, *Arthur1*, *TC1_XL*, *TC1_DM*, *hAT-4_Crp*, *hAT-14_Crp*, *Zaphod*, *Zaphod2* and *Zaphod3*), the copy number was too low, and we could not generate a reliable alignment. In such cases, we directly utilized the TSDs reported in the literatures¹² for the corresponding TE family (if available) or superfamily. Furthermore, we identified TIRs by aligning the 5' and 3' terminal 500-bp sequences of each TE. We defined well-aligned regions covering the starting nucleotide of the consensus as TIRs. Occasional mismatches near the terminal regions were observed. To accommodate this, specific parameters were implemented for allowing mismatches in TIRs: 1) up to two mismatches were permitted within a continuous stretch of five nucleotides; 2) in cases of three or four mismatches, a continuous stretch of 10 nucleotides was further examined. The latter criterion mainly applied to the *Tc1* family TEs due to their potential to harbor long TIRs (> 100 bp, Figure 3A) and exhibit a greater number of mismatches. To further control the redundancy where multiple species shared an ancient TE emerging in their common ancestor, we selected the species with the lowest average percent divergence between TE copies and the corresponding Repbase consensus.

After these analyses, we generated the final dataset including 124 candidate TEs with essential domains, TIRs and TSDs (Table S2A). Notably, among the 20 known active TEs, the species harboring eight TEs (*ZB*, *Passer*, *Tc1*, *Hsmar1*, *TnpDM*, *THAP9*, *Harbinger3_DR*, and *piggyBat*) were included in these 100 species. For these eight TEs, four TEs (*Passer*, *TnpDM*, *THAP9* and *piggyBat*) were not present in Repbase. We excluded *Tc1*, *Hsmar1* and *Harbinger3_DR* to focus on novel TEs. Only *ZB* was included in the 124 candidates because its activity was not reported until 2021.⁵⁹

The reason that we focused on Repbase consensus sequences rather than phylogenetically reconstructed sequences is that the latter strategy could be complicated by TE horizontal transfer.³¹ Certainly, it should be noted that consensus sequences could represent nonautonomous TEs with high copy number, and our heterologous screening would fail in this scenario. However, in previous case studies,^{33,34,86,134,135} a consensus strategy has been routinely used. We thus chose this strategy, considering its simplicity and scalability.

De novo annotation of six active TEs in two fish genomes

We implemented a separate pipeline for two *de novo*-assembled fish genomes without TE annotation. Initially, RepeatModeler v1.0.11¹²⁶ was employed with the default parameters to generate the TE family consensus for each genome. Similar to the previous pipeline, ORFfinder (or GENESCAN for TEs from the *P* and *PIF/Harbinger* superfamilies) and PfamScan were used to identify TEs with ORFs longer than 300 aa and encoding a transposase domain. Only 12 *Tc1/mariner* TEs were retained. However, alignments of two termini indicated a lack of intact TIRs for the majority of consensus TEs. Therefore, on the basis of the TE copies most similar to the consensus, we manually extended the TIRs by searching for the longest complementarily aligned flanking sequences associated with “TA..TA” TSDs (canonical TSD of the *Tc1/mariner* superfamily²⁵; Figure S3F). After removing redundancy between TEs as reflected by the same transposase and TIRs, we finally retained six unique TEs (Table S2B). Based on these six reconstructed consensus sequences, we reran RepeatMasker to search the genome and identified potentially full-length TE copies, as we did for UCSC-annotated genomes. Notably, these TEs are named based on the Repbase convention, where the TE family name is followed by a numeric ID and the initials of the species' Latin name.

Phylogenetic modeling of activity-related factors

We first reconstructed the phylogenetic tree of 130 candidate TEs together with 20 known TEs (Table S1). A total of 149 sequences were involved since the recently discovered *Tc1_8B_DR* or *ZB*⁵⁹ was included in our candidate list. We aligned the amino acid sequences of the relatively conserved catalytic DDE/D domain using MAFFT. We then generated the phylogenetic tree using IQ-TREE v2.1.4-beta.¹¹⁹ To accelerate the tree building process, we used the following parameters according to the authors' recommendation: -m MFP -bb 1000 -bnni -redo -T AUTO. We selected IS30 as the outgroup to root the tree since bacterial IS elements are distantly related to eukaryotic DNA TEs.^{11,136} We visualized the resulting phylogenetic tree using iTOL v6.¹²⁰

We then performed phylogenetic logistic regression analysis by utilizing the R package *phylolm* v2.6¹³⁷ and following the approach proposed by.¹³⁸ Initially, we performed AIC-based stepwise model selection starting from the full model including the following factors: the copy number of potentially full-length autonomous TEs, superfamily and family (in case of *Tc1/mariner* and *hAT*), average percent divergence relative to the corresponding consensus sequence and species divergence time (retrieved from TimeTree¹³⁹) relative to human. This process led us to identify the best model, which utilized copy number as the sole predictor. However, importantly, other models with a Δ AIC less than 2 were also plausible (equivalent to a probability lower than 2.7-fold compared to the reported best model), indicating some uncertainty regarding the best model. Consequently, we conducted model averaging by considering all possible combinations of the four variables. Individual models were built and weighted based on Δ AIC.¹⁴⁰ The relative importance of each variable was determined by summing its weight across all models containing that particular variable.

Identification of nonautonomous TE copies

During the aforementioned reannotation process (searching genome with consensus sequences), we also identified nonautonomous TE copies. In brief, we extracted candidate nonautonomous TEs as two adjacent TE fragment hits with intervals shorter than the length of the consensus, *i.e.*, one broken RepeatMasker alignment potentially caused by the internal deletion of an autonomous TE. To ensure genuine TE insertions, we applied three filters: 1) the two fragments possessed intact termini matching the 5' and 3' ends of the consensus; 2) we restricted the length of 'N's (unknown bases) between adjacent TE fragments to less than 20% of the total TE length to avoid identifying fragmentations caused by assembly gaps; and 3) TSDs are present in the immediate flanking region, indicating a single TE insertion. The retained copies, excluding the previously defined autonomous copies, were considered as nonautonomous copies. These copies were then plotted in R as a scatter plot, illustrating their breakpoint positions relative to the consensus sequence within the two fragments (Figure S2A; Mendeley Data: <https://doi.org/10.17632/bycpd2tcxw.1>). In addition, according to the position of the breakpoint, we inferred whether miniature inverted repeat transposable elements (MITEs) have complete TIRs.

Classification of TE evolutionary dynamics in terms of MITEs

We conducted an analysis of the breakpoints of nonautonomous copies and determined the number of nonautonomous copies with breakpoints occurring at the same position or within a 50-bp vicinity to account for issues such as alignment uncertainty. Nonautonomous elements with at least three copies at the same position were defined as a MITE. For active DNA TEs, the distribution of MITEs could be classified into three scenarios or models: the majority of nonautonomous copies are MITEs; TEs lack MITEs, and the nonautonomous copies demonstrate random internal deletions, leading to length heterogeneity; and in a hybrid scenario, one or more MITEs are present, but the majority exhibit random deletions.

Analyses of direct repeats in the *Tc1/mariner* superfamily

Since direct repeats (DRs) could be situated in TIRs or the flanking spacers (Figure 1A),¹¹ we searched both regions. We developed in-house Perl scripts to search DRs with a minimal repeat of 5 bp in order to tolerate mismatches in DRs and increase the sensitivity. We required the simultaneous presence of DRs on both sides. Since the search could return more than one type of DR candidate for one TE, we manually curated the results by prioritizing paired longer DRs with a symmetrical distribution on two sides (Table S4). As shown in Figure 3A, we classified TEs into four types based on the length of TIRs, the number of DRs, and the relative positions of DRs and TIRs.

For Figure 3A, the phylogenetic logistic regression¹³⁸ comparing TIR-DR types of active and inactive *Tc1* elements was conducted using the R package *phylolm* v2.6. For Figure 3B, the phylogenetic ANOVA¹⁴¹ predicting activity quantitatively for active *Tc1* elements was performed using the R package *geiger* v2.0.11.¹⁴²

Identification of critical residues associated with transposition activity

We identified conserved (unchanged) amino acids of active *Tc1* transposases and assessed whether they affected transposition activity. Specifically, different from Figure 2A with only deeply conserved catalytic DDE/D domain alignment, we took advantage of similarity of *Tc1* members and performed full sequence alignment of 40 *Tc1* TEs in this study, the well-studied SB100X and *Tc1* elements known to be active in human cells (*Minos*, *Tc1*, *Passport* and *FP*; Figure 2A). By processing the multiple sequence alignment (Mendeley Data: <https://doi.org/10.17632/ktfwtk6k3f.1>) with an in-house script, we identified 23 residues shared by all active transposases. To evaluate whether these residues affect transposition, we first conducted a literature review and found that eight residues had been experimentally shown to be detrimental to transposase activity when mutated. These included G59 of the GRPR-like motif in *Rag-1*¹⁴³ and DDE or SPDL motifs in *Mos1*.^{144,145} For the remaining 15 novel residues, we selected phylogenetically representative

SB100X and *Tc1-2_ST* (Figure 2A) as targets to evaluate whether the mutations affect transposition activity. Among the 15 residues, R131, W195, Y218, W282, and R326 were mutated to the corresponding residues found in inactive TEs. The others, such as residues corresponding to multiple derived states, were mutated to alanine, the smallest residue. Subsequently, we quantified the transposition efficiency of TEs with these respective mutations in HEK293T cells, and the activities of the mutants were further normalized to their respective wild-type transposases.

Analogously, we analyzed *hAT* transposases (Mendeley Data: <https://doi.org/10.17632/b54z6nv7t.1>) encoded by 43 *hAT* elements from this study, *Hermes* with a crystal structure,⁷² and five *hAT* transposases previously known to be active in human cells (*Tol1*, *TcBuster*, *SPIN*, *Tol2* and *Ac*; Figure 2A). We identified 15 conserved residues in all 18 active *hAT* transposases, including the conserved CCHH motif in the BED domain^{61,146} and the DDE motif in the catalytic domain.¹⁵ W319 guided hairpin formation when the transferred DNA was cleaved from the original location in the transposition process of *Hermes*.^{72,147} To verify the function of the remaining seven novel residues, we selected phylogenetically diverged *hAT-7-PM* and *HAT1_AG* as mutation targets and assessed their function in HEK293T cells. The transposition activities of the mutants were also normalized to their respective wild-type transposases.

Additionally, we employed two alternative tools, PolyPhen-2 v2.2.3¹⁰² and ESM1b,¹⁰³ for identification of critical residues. Although originally developed to predict the effect of amino acid substitutions, we utilized them to infer the importance of residues by considering residues associated with damaging mutations as critical. Specifically, PolyPhen-2 utilizes homology search across nucleotide, protein, and structural databases to extract features for subsequent machine learning. The default HumDiv was used as the probabilistic classifier, generating a Naive Bayes probabilistic score ranging from 0 to 1. Mutations with scores > 0.85 are typically classified as "probably damaging". Conversely, ESM1b is a protein language model trained on approximately 250 million protein sequences. We derived a pseudo-log-likelihood ratio (PLLR) comparing wild-type protein with potential missense mutations based on ESM1b model weights. Variants with lower PLLR scores are predicted to have a more deleterious effect. In our analysis, a PLLR threshold of -5 was utilized to distinguish deleterious variants. To be conservative, we applied two levels of filtering. For each amino acid within each transposase, we assessed mutations to the other 19 amino acid types using both tools. A critical amino acid is defined as one where mutations to any other amino acids are deemed deleterious (score > 0.85 for PolyPhen-2 or PLLR < -5 for ESM1b). Subsequently, by mapping these amino acids to the coordinates of SB or *Hermes* based on multiple sequence alignment, a critical residue was designated if all homologous amino acids were deemed critical.

Protein structural modeling

The protein structures were predicted using AlphaFold v2.0.0.¹¹² Structure images were generated in PyMOL v 2.5.5 (<https://pymol.org/>). The focus was on the structures of the catalytic domains, as the nonconserved domains could result in low confidence in structure prediction. For the transposase-TIR complexes of *hAT-7-PM* and *HAT1_AG*, modeling was performed using ChimeraX 1.4,¹¹⁵ with the crystal structure of the *Hermes* transposase-TIR complex as the template.⁷² Structure alignments of SB100X and *Tc1-2_ST*, as well as *Hermes* and *hAT-7-PM* and *HAT1_AG*, were also conducted in ChimeraX, along with the calculation of hydrogen bonds. The overlapping structures of SB100X and *Tc1-2_ST* can be found in "Mendeley Data: <https://doi.org/10.17632/bz7kh3cgwp.1>", while the overlapping complexes of *Hermes*, *hAT-7-PM* and *HAT1-AG* are shown in "Mendeley Data: <https://doi.org/10.17632/y4xtj2yn7m.1>".

Plasmid construction

Plasmids for transposition screening in HEK293T and HeLa cell lines

In Tables S2 and S3, we have provided the amino acid sequences of transposase (Tnp) or Myb-like protein (myb for *PIF/Harbinger*), transposon fragments (5' and 3' transposon sequences after removing the internal Tnp), and TSDs. These sequences were synthesized by BGI (BGI, Suzhou, China). The Tnp sequences underwent codon optimization to match the human codon usage (Table S3A). With EcoRI and NotI restriction enzyme sites,¹¹³ the codon-optimized Tnp sequences replaced the hyPBBase ORF in the pCMV-hyPBBase vector (hyperactive *piggyBac*³²; Figure 1D). These vectors served as helper plasmids (referred to as pCMV-Tnp) in the experiments (Table S3A). The donor plasmid consists of the PGK-puro/eGFP-bGH poly(A) cassette flanked by the transposon fragments and TSD (Table S3B). It was cloned into the general-purpose cloning vector pMV or pQLL (provided by BGI) with the same construction procedure as the helper plasmids.

Plasmids for transposase mutant experiments

The helper plasmids were constructed based on the wild-type transposase plasmid. For Figure 4, the expression plasmids used were pCMV, while for Figure S7, pICOZ expression plasmids were utilized (Table S3C).

Plasmids for cargo capacity evaluation

Helper and donor plasmids were also employed in the experiments. The helper plasmids or pCMV-Tnp remained unchanged, while the donor plasmids were categorized into three types based on the length of the carrier genes (Table S3C). The first type encompassed a 2-kb cassette ([PGK-puro/eGFP-bGH poly(A)]) positioned between two transposon fragments, identical to that used in the previous transposition activity detection experiments (Figure 1D; Table S3B). The second type accommodated a 5-kb cargo gene with an additional 3-kb segment ([EF1between twV40 poly(A)]) appended to the first type. The third type featured a cargo gene of up to 10 kb, with the *HBB* locus (5 kb) positioned at the 3' end of the [EF1 with an additional 3-kb segment ([EF1between

twV40 poly(A)) appended to the first type. The third constructed with pMV vectors, with the respective cargo genes (2-kb, 5-kb, and 10-kb) inserted between the transposon fragments (Figure 1D).

Plasmids used in primary T cells

To be consistent with the previous literatures on CAR-T based gene therapy,^{92,129} we replaced the [PGK-puro/eGFP-bGH poly(A)] cassette in pMV or pQLL vectors with the [EF1α-tween twV40 poly(A)] appended to the first type. The third constructed with Table S3C). Helper plasmids (pCMV-Tnp vectors), with a length of approximately 5.7 kb, were still utilized as before. For TE CAR-T cells, the [EF1α-CAR-bGH poly(A)] cassette was cloned into transposon fragments, with TSD sequences flanking the fragments in donor plasmids. To mitigate cell toxicity caused by redundant bacterial sequences in plasmids and improve efficiency in transferring larger DNA fragments, the donor and helper plasmids were further minimized using the pCOZ vector (1.1 kb) to replace the pMV (2.1 kb) and pCMV (3.5 kb) vectors, respectively (Table S3C). For the generation of viral CAR-T cells, lentiviral vectors were utilized, including the Fuw-EF1α-CAR plasmid and pRRLSIN-EF1α-CAR plasmid, along with packaging plasmids (pMD2.G and psPAX2, Table S3C). Specifically, the Fuw-EF1α CAR-T cells, lentiviral vectors were utilized,¹²⁸ while the pRRLSIN-EF1α Fuw-E¹⁴⁸ was generously provided by Dr. Weidong Han (Department of Bio-therapeutic, Department of Molecular & Immunology, Chinese PLA General Hospital, Beijing, China). CD19 and HER1 CARs were cloned into these two plasmids, respectively.

Transposition assay

Transposition assay in HEK293T cells

In the initial screening experiment, HEK293T and HeLa cell lines were employed to assess the transposition activity of the candidates. For each well of a 24-well plate, 1.2×10^5 HEK293T cells were seeded one day before transfection. For transfection, 300 ng of plasmids was used (200 ng helper plasmid and 100 ng donor plasmid), along with 1 μ L Lipofectamine® 2000 reagent (Thermo Fisher Scientific) following the manufacturer's recommendations. To serve as a transposition control, pRS313 was co-transfected with the donor plasmids. Each TE candidate was assayed with two biological replicates. After 48 hours of transfection, the proportion of GFP-positive cells was determined using flow cytometry, representing the transfection efficiency to ensure the successful entry and expression of plasmids within the cells. One percent of the transfected cells were further replated in a 10 cm plate and cultured in selection medium containing DMEM supplemented with 0.5 μ g/mL puromycin. After 10 days of puromycin selection, the cells were fixed with 4% cold paraformaldehyde and stained with 0.2% methylene blue. The blue colonies were then quantified using ImageJ v1.48. Active TEs were identified as those with a colony count 1.5 times higher in the group transfected with both helper and donor plasmids compared to the group transfected without the helper plasmid.

Transposition assay in HeLa cells

The procedure in HeLa cells largely mirrored that in HEK293T cells. The differences included initially seeding 7.5×10^4 HeLa cells prior to transfection, diluting the transfected HeLa cells by 10%, and culturing the diluted transfected cells in selection medium for 14 days. These changes were made because HeLa cells are larger and grow more slowly than HEK293T cells.

Transposition assay in Neuro-2a cells

The procedure in Neuro-2a cells closely resembled that of HEK293T cells, with the only notable difference being the initial seeding number, adjusted to 1.0×10^5 cells.

Transposition assay in K562 cells

To assess the transposition activity in suspension cells (e.g., K562), the transfection mixture, identical to that used in adherent cell lines (HEK293T, HeLa and Neuro-2a), was initially added to each well of a 24-well plate, followed by the addition of 1.0×10^5 K562 cells in 300 μ L medium. After 12 hours, the culture medium was added to each well to achieve a final volume of 1 mL. The transfected cells were passaged every three days until the proportion of GFP-positive cells in the group transfected with only donor plasmids decreased to nearly 0% after 12 days. At this moment, the proportion of GFP-positive cells in the group transfected with both donor and helper plasmids on the 12th day represents the transposition efficiency.

Transposition assay in T cells

Given that T cells are also suspension cells, we followed a procedure similar to that employed for K562 cells. Specifically, to determine the transposition efficiency of the four novel active TEs and SB100X in primary T cells, we introduced pCMV-Tnp (800 ng) and pMV or pQLL-Transposon Fragment [EF1α-CopGFP-SV40 poly(A), 800 ng] into 1×10^6 resting CD3⁺ T cells using 20 μ L Nucleovette™ strips and the EO-115 program of the Amaxa™ 4D-Nucleofector following the manufacturer's instructions (Lonza, Cologne, Germany). For control groups, only donor plasmids (pMV or pQLL-Transposon Fragment [EF1α-CopGFP-SV40 poly(A)]) were electroporated to exclude transient expression and evaluate transposition efficiency more precisely. Five hours after transfection, the transfected T cells were activated using anti-CD3/CD28 Dynabeads (Thermo Fisher). Cell viability was assessed 48 hours post-electroporation through trypan blue staining. Subsequently, the proportion of GFP-positive cells was measured with flow cytometry every three days. The proportion on the 12th day represents transposition activity.

Insertion site analyses

Generation and sequencing of integration-site mapping libraries

We conducted flow cytometry sorting on the 12th day post-transfection to isolate GFP-positive K562 cells, indicative of stable transgene integration through transposition events. Conversely, due to significant primary T cell loss during sorting, unsorted T cells were directly utilized, with reliance solely on subsequent PCR assays for insertion site recovery.

The genomic DNA was extracted from the cells using the DNeasy Blood & Tissue Kit (Qiagen) following the manufacturer's instructions. The DNA was then sonicated in Snap-Cap microTUBEs (Covaris) using a Covaris S220 ultra-solicator device under standard conditions (peak incident power 105 W, duty factor 5%, cycles per burst 200, treatment time 80 s) to achieve an average fragment size of 500 bp. The sonicated DNA was purified using 1.5 x VAHTS® DNA Clean beads (Vazyme), and fragments ranging from 200 to 500 bp were selectively obtained and purified with a 1% agarose gel. Then, the DNA underwent damage repair, end blunting, and 3'-A-tailing using the VAHTS® Universal Pro DNA Library Prep Kit (Vazyme). The DNA was further cleaned using 1.5 x VAHTS® DNA Clean beads and subsequently ligated with a Y-adaptor,¹⁴⁹ generated by annealing the Illumina common oligo with the barcode adapter corresponding to each TE (Table S5). The ligation mixture was further purified using 1.2 x VAHTS® DNA Clean beads.

The eluted DNA was subjected to the first round of PCR using the P5-1 primer specific to the linker and TE/gene specific primers (GSP1) targeting the left end sequence of TIR. The PCR employed PrimeSTAR® HS DNA Polymerase (Takara) with the following conditions: 98°C for 5 min; 15 cycles of [98°C for 30 s, 70°C (-1°C/cycle) for 2 min, 72°C for 30 s]; 10 cycles of (98°C for 30 s, 55°C for 1 min, 72°C for 30 s); 72°C for 5 min; 4°C hold. The PCR products were purified using 1.2x VAHTS® DNA Clean beads and eluted in 15 µL of 1 x Tris-EDTA buffer, pH 8.0, then used as a template for the second round of PCR with the P5-2 nested primer for the linker and GSP2 primers combined with specific primers for TIR and P7 adapter. The conditions for the second PCR were identical to those of the first round. The final PCR products were purified using 1.2x VAHTS® DNA Clean beads and eluted in 20 µL of 1 x Tris-EDTA buffer with a pH of 8.0. Finally, the samples were sequenced on the NovaSeq 6000 platform (Novogene) with the 150-bp paired-end run settings.

Insertion-site identification

Clean data were obtained by selecting sequencing reads that passed the quality filter and ended with primer sequences corresponding to TEs. Cutadapt v3.4¹¹⁶ with the default parameters was used to trim sequences containing TIR and adapters. The remaining reads were mapped to the human genome (UCSC hg38) using Novoalign v3.09.04 (Novocraft Technologies, <http://novocraft.com/>) given its accuracy.¹⁵⁰ PCR duplicate reads were removed using the Picard MarkDuplicates tool v2.26.1 (Broad Institute, <https://broadinstitute.github.io/picard/>). To determine the integration sites of TEs in K562 cells, the following procedure was implemented: 1) uniquely mapped and properly paired reads with no more than 5-bp indels or mismatches in total were retained, disallowing soft clipping at the ends of TEs; 2) within a 100-bp bin, only positions supported by at least two reads were considered as true integration sites, while others were unreliable and discarded.

For *Tigger4*, additional filtering was applied to distinguish new integration sites from native ones preexisting in the human genome. Initially, the hg38 genome was searched for *Tigger4* fragments using RepeatMasker. Hits precisely aligned with the left or right ends of *Tigger4* consensus sequences were selected, and the corresponding genomic loci were extracted. Insertion sites identified from the sequencing data were then compared with these loci using bedtools v2.30.0.¹¹⁴ Any sites matching these locations were excluded as primer mis-targeting events.

The coordinates of the insertion sites for various TEs can be found in "Mendeley Data: <https://doi.org/10.17632/2bm6x6235k7.1>".

Analyses of insertion-site preferences

We focused on four aspects: location at/in genes, location at transcription start sites (TSSs), chromatin states, and safe harbors.

First, we obtained gene annotations for the human genome (hg38) from the UCSC Genome Browser (track ncbiRefSeq). The distance between the integration site and the nearest gene was calculated and categorized into 10 kb intervals. Similarly, the distance to TSSs was processed using 500-bp intervals. Chromatin state and histone mark annotations for K562 cells were downloaded from the Roadmap Epigenomics project,⁷⁹ including 15 ChromHMM states and broad histone domains (epigenome ID: E123). Coordinates of these annotations were converted from hg19 to hg38 using the UCSC LiftOver tool. The integration sites were subsequently examined for overlap with chromatin states or histone marks using bedtools.

We then identified genomic safe harbors, defined as regions in the human genome capable of accommodating the expression of newly integrated DNA without strong adverse effects on the host.⁸¹ Safe harbors needed to meet specific criteria, including a distance of > 50 kb from TSSs, > 300 kb from cancer genes, and > 300 kb from microRNA genes, location outside of genes, and location outside of conserved regions.⁸¹ The cancer gene list was obtained from the oncoKB database,¹⁵¹ while TSS annotation, microRNA gene annotations and conserved regions were from UCSC Genome Browser (track ncbiRefSeq, wgRna and phastConsElements30-way, respectively). We assessed whether the integration sites overlapped with the identified safe harbors using bedtools.

To establish a random control dataset, we generated simulated reads with realistic error profiles using pIRS v2.0.2.¹²⁴ Paired reads with a length of 100 bp and an insert size of 300 ± 50 bp (mean $\pm$ SD) were simulated for the hg38 genome at a coverage level of 0.05. The depth surpasses the actual values, guaranteeing ample genome coverage and facilitating subsequent fold enrichment calculations. These simulated reads were mapped back to the genome, and the resulting insertion sites served as control. We counted the number of integration sites and control sites within the four sets of genomic features mentioned earlier. For each feature, we calculated the odds ratio of integration sites compared to control sites. TEs were heuristically clustered using complete linkage based on the Euclidean distance of their log odds ratios in R 4.1.0. Poisson bootstrapping was performed on all integration sites of each TE, and the same clustering method was applied to each bootstrap sample. Confidence levels were derived from 1,000 bootstrap clustering trees using the ape 5.5 package. Branches with bootstrap confidence scores lower than 60 are collapsed. Note that the clustering's resolution power appears insufficient as the random control group remains indistinguishable from *Tc1* elements (Figure 5A).

Analysis of TSD motifs

To identify the target motifs of various TEs in their native genomes, we extracted the 20-bp flanking sequences of the putative autonomous TEs given the known motif size (< 20 bp, Table S2). Additionally, to investigate the integration preferences of TEs in K562 cells, we extracted 20-bp sequences surrounding the integration sites in the human genome. Using WebLogo 3.7.8,¹²⁷ we constructed sequence logos and position weight matrices (PWMs) for both the native and human genomes based on these extracted sequences. This allowed us to visualize the motifs and their sequence patterns.

To assess the similarity between the integration sites in the native genomes and K562 cells, we incorporated the PWMs built from the K562 insertion sites into the HOCOMOCO Human v11 FULL motif database (downloaded from MEME Suite¹²²). Subsequently, we utilized the MEME Tomtom tool 5.3.0 to search for the corresponding native genome motif within the merged database. The resulting *P*-value, indicative of the match between the two motifs, was employed as a measure of statistical similarity.

Chromatin accessibility analysis during CAR-T cell exhaustion

We collected public Assay for Transposase-Accessible Chromatin with high-throughput sequencing (ATAC-seq) data from the NCBI GEO Database (Accession GSE160139).¹⁵² Unstimulated and exhausted CAR-T cell samples were included in the analysis. Utilizing the ENCODE ATAC-seq pipeline v2.2.2¹¹⁷ with default parameters, raw FASTQ files from four replicates per condition were processed to identify optimal overlap peaks, ensuring enhanced sensitivity in downstream analyses. Subsequently, mapped insertion sites in T cells were examined for overlap with these ATAC-seq peaks using bedtools.

Cargo size tolerance of the top active TEs in the three superfamilies

Similar to previous efforts,¹⁵³ we constructed donor plasmids containing genes of three different lengths: 2 kb, 5 kb, and 10 kb (see also “Plasmid construction” section, Tables S3C, S4, and S5). The experimental procedure followed the same steps as in the transposition assay in HEK293T cells. To account for variable transfection efficiency possibly caused by distinct plasmid size, we calculated transposition efficiency as follows: transposition efficiency = puromycin resistance colony number / (cells number on the 2nd day after transfection × transfection efficiency).

Overproduction inhibition assays

Since the overproduction inhibition of SB100X has been studied in HeLa cells,¹⁵⁴ we herein used this cell line. The procedure is similar to that described for the “Transposition assay in HeLa cells”. The donor plasmids utilized were pMV or pQLL-transposon fragment [pGK-puro/eGFP], while the helper plasmid employed was pCMV-Tnp. Throughout the experiment, the transfection dose of the donor plasmid remained constant at 100 ng. Following a previously reported method,¹⁵³ the ratio of helper plasmids was gradually adjusted from 0.125:1 to 2:1 relative to the donor plasmid, resulting in helper plasmid doses ranging from 12.5 ng to 200 ng. To ensure consistent DNA transfection mass across all groups, an unrelated plasmid commonly employed in yeast studies (pRS313,¹⁵⁵ Table S3C) was utilized for complementation. This approach maintained a stable total amount of transfected DNA, specifically 300 ng in the 2:1 group. Transposition activity was once again quantified using colony counts. To evaluate whether TEs exhibited overproduction inhibition, we normalized the data by considering the highest observed transposition activity for each TE. This normalization facilitated a comparative assessment of transposition activities across different doses.

CAR-T-related experiments

Generation of lentiviral CAR-T cells

Lentiviral CAR-T cells were produced as previously described.¹²⁹ In brief, 12 µg FUW/pRRLSIN-CAR plasmid, 6 µg pMD2.G, and 6 µg psPAX2 packing plasmid DNA were co-transfected into HEK293T cells using Lipofectamine 3000 (Thermo Fisher Scientific) when HEK293T cells were 95% confluent. The virus supernatant was harvested at 48 h and 72 h post-transfection and then concentrated by ultracentrifugation (Millipore) at 4000 rpm for 1 hour at 4°C. The concentrated virus was aliquoted and stored at -80°C for future applications. Isolated T cells were first activated using CD3/CD28 beads, and 24 hours later, 1 × 10⁶ T cells were resuspended in 1 mL of T cell medium and transduced with 50 µL of concentrated virus. Subsequently, 1 µL of polybrene was added to increase the transduction efficiency. Notably, the proportion of anti-CD19 CD28 (the co-stimulatory domain) CAR-positive cells in the LV-CD19 CD28 group is markedly low, with a positive rate of less than 4% observed on the 6th day post lentiviral infection. Consequently, cells from this group underwent a round of enrichment via flow cytometry sorting.

Generation of TE CAR-T cells

For each sample, 5 × 10⁶ T cells (activated by human CD3/28 Dynabeads for 48–72 hrs) were centrifuged at 300 × g for 5 min, and then washed with 37°C warmed Opti-MEM (GIBCO) one time. Given the rapid change of electroporation techniques,^{109,156,157} we employed three systems—Maxcyte STX, Lonza 4D, and B1mix—to deliver TE plasmids, each exhibiting moderately varied performance outcomes. For Maxcyte STX, T cells were resuspended in 50 µL Opti-MEM containing 5 µg helper and 5 µg donor plasmids. Cells were mixed gently and then transferred into electroporation cups (OC-100). Program-Expanded T cell 3 was used for human activated T cells. For Lonza 4D, T cells were resuspended in 100 µL transfection buffer containing 5 µg helper and 5 µg donor plasmids. The P3 Primary Cell 4D-Nucleofector X Kit (V4XP-3024, Lonza), and program EO-115 were used for human stimulated T cell electroporation. For B1mix, T cells were resuspended in 100 µL B1mix buffer containing 5 µg helper and 5 µg donor plasmids. Electroporation cups (V4XP-3024, Lonza) and program EO-138 were used for human stimulated T cell electroporation.

Using either platform, immediately after electroporation, the cells were carefully removed from the cups and transferred to 6-well plates, and the cups were rinsed with the same volume of Opti-MEM. After resting for 20 min at 37°C with 5% CO₂ in a humidified incubator, the cells were resuspended in complete culture media at 2.5×10^6 cells/mL. The percentage of CAR-positive cells was calculated 2 days and 12 days after electroporation with Alexa Fluor 647 AffiniPure Goat Anti-Mouse IgG, F(ab')₂ Fragment Specific (for CD19 CAR-T cells) or Alexa Fluor 647 AffiniPure F(ab')₂ Fragment Goat Anti-Human IgG (H+L) (for HER1 CAR-T cells) by flow cytometry (BD Fortessa).

Luciferase-based cytotoxicity assay

K562-CD19-luciferase cells and H266-luciferase cells (H266-luci) were used to test the cytotoxic ability of CD19 CAR-T and HER1 CAR-T cells, respectively. Briefly, tumor cells were resuspended in RPMI 1640 medium at a density of 1×10^5 cells/mL and then seeded in white opaque 96-well plates at 100 μ L per well. Effector cells were added to each well corresponding to a specific E:T ratio, resulting in a final volume of 200 μ L/well. The plates were incubated at 37°C in 5% CO₂ for the required time. Four E:T ratios were tested for CD19 CAR-T cells. Cells were co-incubated for 96 hours when the E:T ratio was 0.125:1 or 0.25:1, and for 48 hours when the E:T ratio was 0.5:1 or 1:1. At the end of co-incubation, 10 μ L of Steady-Glo luciferase substrate (Promega) was added, and the reaction was carried out for 5 minutes at room temperature. PerkinElmer VICTOR X3 was used to measure luminescence and the percentage of specific lysis was calculated using the equation: % killing = $100 - [(RLU \text{ from well with effector and target cell coculture}) / (RLU \text{ from well with only target cells}) \times 100]$. Notably, according to this formula, the resulting killing percentage could potentially be below zero.

Multiple rounds of tumor cell challenge

In the initial round, CAR-T cells were co-incubated with corresponding target cells at an E:T ratio of 1:1. In each subsequent round, a consistent number of fresh tumor cells, identical to the initial round, was added. Following a 48-hour co-incubation period in each round, a small aliquot of cells with equivalent volumes from each group was harvested to assess cytotoxicity efficiency and the proportion of CAR-positive cells. This process continued until the CAR-T cells cytotoxicity capability was significantly reduced. The percentage of specific lysis in every round was measured by luciferase assay using PerkinElmer VICTOR X3. CD19 CAR-T cells with or without four rounds of tumor co-incubation were collected to analyze cell subtypes. PE anti-human CD4 and APC anti-human CD8a antibodies were used to analyze the ratio change in CD4⁺ and CD8⁺ T cells. PE anti-human CD45RO and APC anti-human CCR7 antibodies were used to analyze the ratio change of CCR7-CD45RO⁻ (effector), CCR7+CD45RO⁻ (naive), CCR7+CD45RO⁺ (central memory, CM), and CCR7-CD45RO⁺ (effector memory, EM) subtypes. The percentage of CAR positive cells was detected after every round of tumor co-incubation by co-staining with Pacific Blue™ Mouse Anti-Human CD3 antibody and PE-Labeled Human CD19 (20-291) Protein (for CD19 CAR-T cells) or and FITC-Labeled Human EGFR Protein (for HER1 CAR-T cells).

Evaluation of CAR-T cell function using mouse models

To evaluate the efficacy of CD19 CAR-T cells, we established a Raji-luci cell line-derived xenograft (CDX) model. Five- to six-week-old female NOD-Prkdcscid Il2rgnull (NPG) mice (one kind of immunodeficient mouse, VitalStar, Beijing, China) were injected intraperitoneally with 3×10^5 or 5×10^5 Raji-luci cells. After 3 days, tumor engraftment was evaluated by serial biophotonic imaging using the IVIS® Spectrum (Perkin Elmer). Mice with comparable tumor burden were divided into the following 5 groups: PBS buffer, T cells, LV-CD19 CAR-T, SB-CD19 CAR-T and MAG-CD19 CAR-T groups. A total of 2×10^6 or 1.5×10^6 cells suspended in 100 μ L PBS buffer were administered by intravenous injection. Tumor burden was evaluated weekly by serial biophotonic imaging, and peripheral blood samples were collected to analyze the proportion of human T cells by flow cytometry. In addition, the survival status of the treated mice was recorded.

To evaluate the efficacy of HER1 CAR-T cells, we established an H266 cell line-derived CDX model by subcutaneous inoculation of 2×10^6 H266 tumor cells into six-week-old NPG mice. When the tumor volume increased to 200–300 mm³, mice were randomly assigned to five groups: PBS buffer, T cells, LV-HER1 CAR-T cells, SB-HER1 CAR-T cells, and MAG-HER1 CAR-T cells. A total of 3×10^6 cells were administered to the mice by intravenous injection. The tumor size was monitored weekly. Peripheral blood samples were collected on the 22nd day to analyze the proportion of human T cells by flow cytometry. On the 52nd day, tumors from mice treated with PBS buffer were dissected into small pieces for tumor re-inoculation of the mice with complete tumor clearance, comprising new mice with re-inoculation but without CAR-T cell injection.

Exploring the mechanisms underlying the heightened cytotoxicity of MAG CAR-T cells

T cells were isolated from four donors and used to generate CD19 or HER1 CAR-T cells utilizing LV or TE vectors. Initially, we implemented an *in vitro* exhaustion model by subjecting CAR-T cells to multiple rounds of tumor challenges.⁹² The experiment was divided into three stages: the unstimulated stage, where modified T cells (CAR-T cells) were cultured for eight days before antigen exposure; the activation stage, where CAR-T cells were stimulated with two rounds of fresh tumor cells; and the potentially exhausted stage, characterized by at least one group of CAR-T cells incapable of effectively clearing tumor cells after at least four rounds of challenges. Notably, in the final stage, it remained plausible that the other two groups still contained a substantial proportion of functional CAR-T cells.

To investigate whether the increased CAR expression in MAG CAR-T cells correlates with heightened tonic signaling, we evaluated the expression of phospho-CD3 ζ in unstimulated CAR-T cells, serving as an indicator of tonic signaling intensity. Furthermore, we quantified the expression levels of Fas, a marker associated with activation-induced cell death (AICD¹⁵⁸), and exhaustion marker genes (PD1, TIM3, LAG3, and CTLA4^{159,160}). Additionally, we assessed the expression of exhaustion marker genes in activated CAR-T cells and potentially exhausted CAR-T cells.

For gene expression quantification in unstimulated CAR-T cells, we utilized fluorescent antibody “Anti-(G4S)n (B02H1) mAb” to detect CAR-T positive cells. We further designed four fluorescent antibody staining combinations: 1) Anti-(G4S)n (B02H1) mAb (FITC Chanel), and recombinant Alexa Fluor® 647 Anti-CD3 ζ (phospho Y83) antibody; 2) Anti-(G4S)n (B02H1) mAb (FITC Chanel) antibody and APC anti-human CD95 (Fas) antibody; 3) Anti-(G4S)n (B02H1) mAb, Brilliant Violet 421™ anti-human CD279 (PD-1) antibody and APC anti-human CD152 (CTLA-4) antibody; 4) Anti-(G4S)n (B02H1) mAb (FITC Chanel), APC anti-human CD366 (Tim-3) antibody, and Brilliant Violet 421™ anti-human CD223 (LAG-3) antibody.

Given the mixed nature of tumor and T cells in activated and exhausted T cell populations, we utilized 7-AAD to exclude dead cells and PE anti-human CD3 antibody staining to isolate T cells. Subsequently, we designed two antibody combinations for analyzing exhaustion marker gene expression: 1) PE anti-human CD3 antibody, Anti-(G4S)n (B02H1) mAb (FITC Chanel), Brilliant Violet 421™ anti-human CD279 (PD-1) antibody and APC anti-human CD152 (CTLA-4) antibody; 2) PE anti-human CD3 antibody, Anti-(G4S)n (B02H1) mAb (FITC Chanel), APC anti-human CD366 (Tim-3) antibody, and Brilliant Violet 421™ anti-human CD223 (LAG-3) antibody. These combinatorial staining ensured the precise detection of gene expression in CAR-positive T cells within the T cell population.

Flow cytometry

Cells were collected and prepared according to the sample loading requirements. Cell analysis was performed in CytoFLEX (Beckman) and BD Fortessa (BD Bioscience). Cell sorting was performed in MoFlo XDP (Beckman) and BD Fusion (BD Bioscience). For the cells that needed to be stained, all operations were performed strictly according to the manufacturers' protocols for the corresponding antibodies. The antibodies used in this study are listed in the “[key resources table](#)”. All data generated by flow cytometry were analyzed using FlowJo software.

A pilot study on optimization of MAG

Increase of activity

Based on the alignment of 23 active *Tc1* transposases (Mendeley Data: <https://doi.org/10.17632/ktfwtk6k3f.1>), we identified semi-conserved residues as those shared by more than 12 active transposases. Among these, MAG encodes 46 specific residues different from the semi-conserved residues. Subsequently, we randomly selected 11 residues from MAG and replaced them with the corresponding semi-conserved residues. The transposition activity of these substitutions was compared with that of the wild-type MAG in HEK293T cells. The screening results showed the increase of activity for M235Q. To validate M235Q, further experiments were conducted in both human HEK293T and mouse Neuro-2a cells.

Mitigation of insertion bias toward genic regions

For SB, three specific residues—H187, P247, and K248—are implicated in target site selection.⁸³ Based on the multiple sequence alignment of 23 active *Tc1* transposases (Mendeley Data: <https://doi.org/10.17632/ktfwtk6k3f.1>), we identified homologous residues in MAG as H181, P240, and K241. Considering their mutational effects on transposition activity and the deviation of insertion sites away from genes observed in SB100X,⁸³ we engineered five MAG transposase mutants: H181A, H181V, P240A, P240R, and K241R. Then, we compared the transposition efficiency of MAG and these mutants in HEK293T and K562 cells. Three mutants (H181A, H181V, and K241R) do not heavily impact its activity. Employing the methodology outlined in the preceding “Insertion site analyses” section, we identified insertion sites for these three mutants in K562 cells and confirmed that they exhibit a weaker genic insertion bias compared to the wild type.

Study approval

All experiments involving human samples and animals were approved by the Ethics Committee of the Institute of Zoology, Chinese Academy of Sciences (IOZ17039, IOZ20170081).

QUANTIFICATION AND STATISTICAL ANALYSIS

The data were presented as box plot, or as mean $\pm$ standard deviation (SD) or standard error of the mean (SEM). Actual data in every replicate were generally overlaid. Depending on the context, the statistical significances were measured by t test, Man-Whitney test, or Fisher's exact test. Unless specified, one-sided tests were performed. Notably, for t tests in [Figures 6](#), [S4](#), [S5](#), [S6B–S6E](#), and [S7E](#), Welch's correction of unequal variance was performed.

All programs were run with default parameters, unless otherwise explained.

Supplemental figures

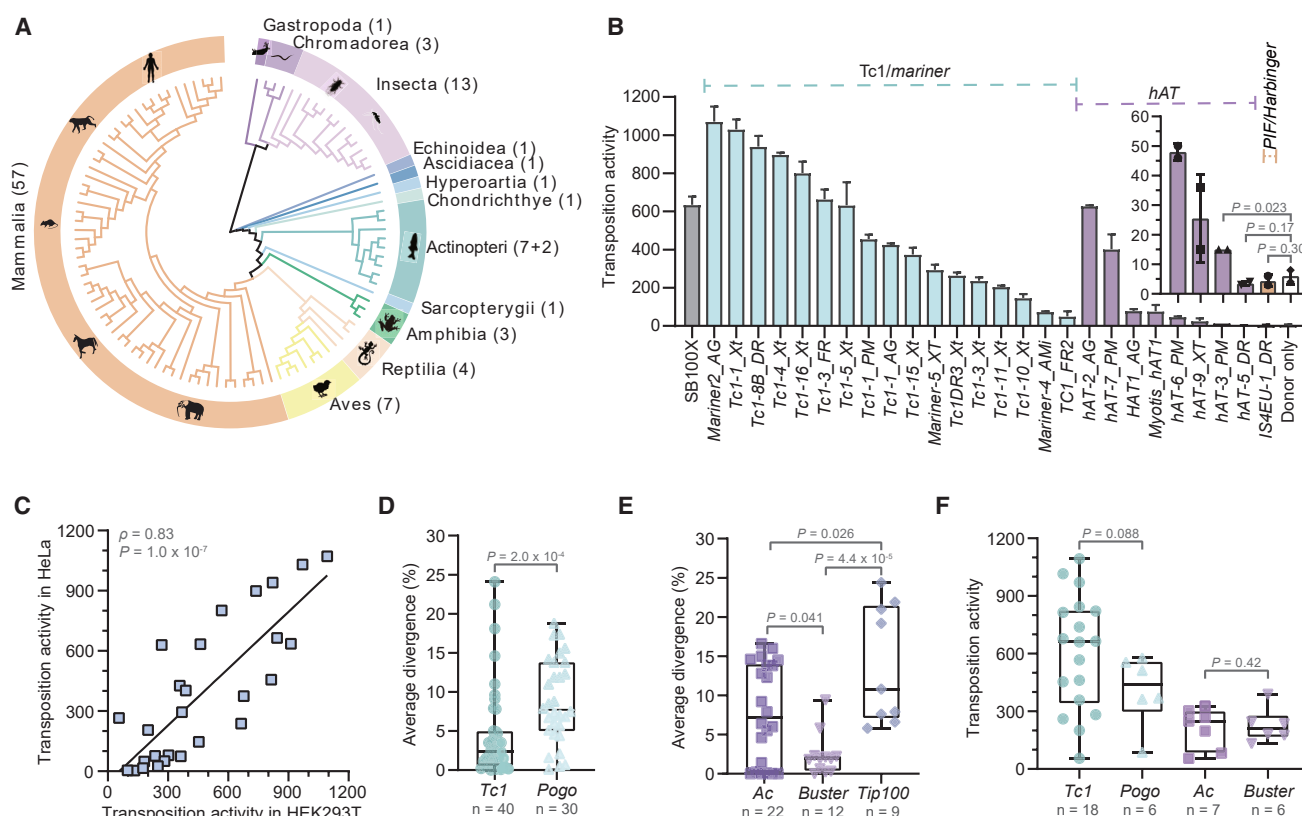


Figure S1. Identification and evolutionary analyses of active DNA TEs, related to Figures 1 and 2

(A) The phylogenetic distribution of 102 animal species. The 102 animals are grouped into 13 classes, which are depicted on the inner branches and outer rings. The number of species assigned to each class is shown within parentheses. Among Actinopteri, seven species possess preexisting TE annotations, while two species undergo *de novo* annotation.

(B) Transposition efficiencies of 26 TEs and SB100X in HeLa cells. The inset on the right is used to better visualize the colony numbers of the last six TEs. Among the 26 TEs, *hAT-5_DR* and *IS4EU-1_DR* are inactive in HeLa cells, whereas all 26 TEs are active in HEK293T cells. Each error bar represents the mean $\pm$ SD of colony numbers on the basis of two biological replicates. For the inset, two actual data points are also overlaid.

(C) The correlation of transposition activity of active TEs between HEK293T and HeLa cells. 26 TEs together with SB100X are plotted.

(D and E) The distribution of the average percent divergence of candidate TEs across the *Tc1* and *Pogo* families (D) or across the *Ac*, *Buster*, and *Tip100* families (E).

(F) The distribution of transposition activity of active TEs across families. Note that there is no active *Tip100* element.

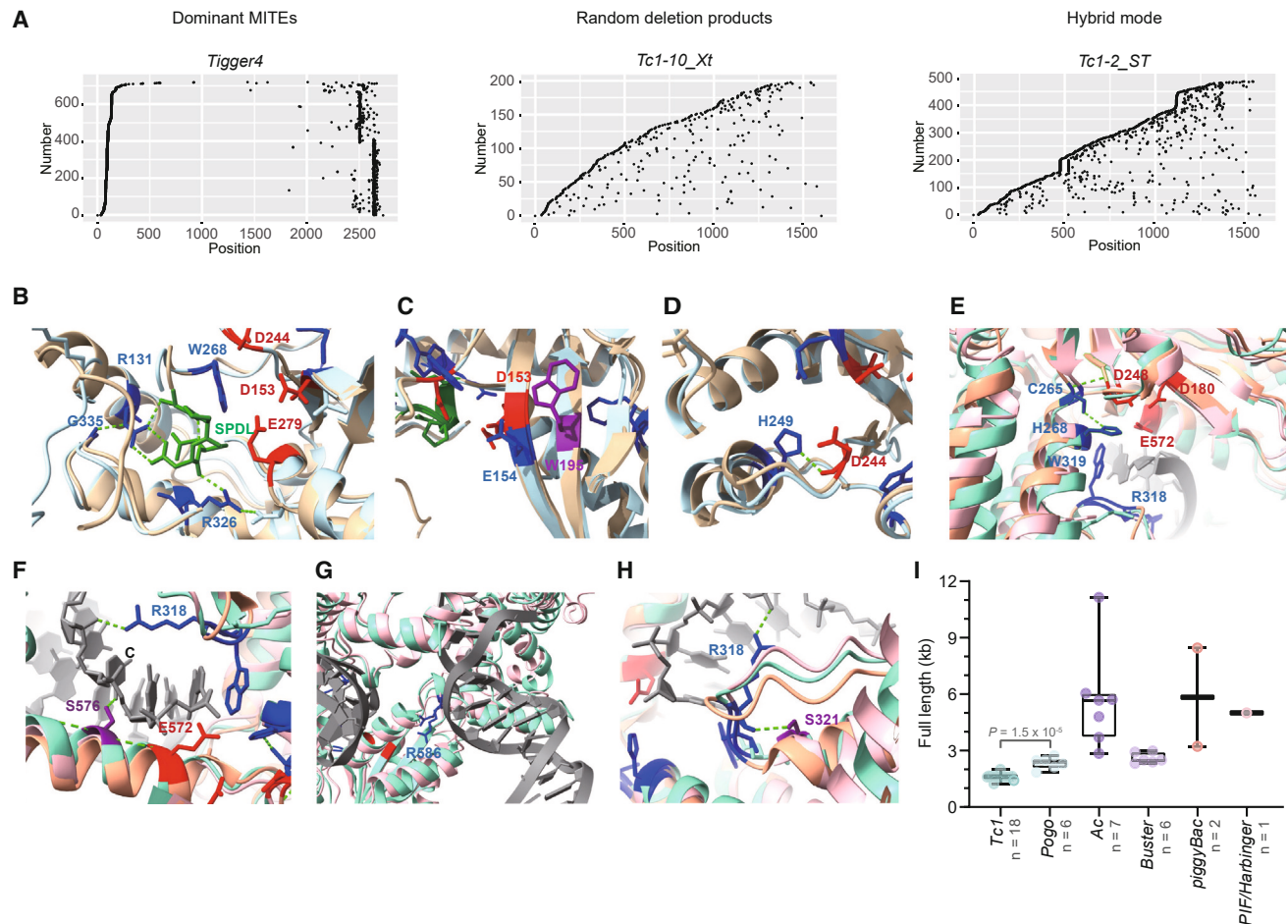


Figure S2. Evaluation of the importance of noncoding sequences and novel conserved residues, related to Figures 3 and 4

(A) The three evolutionary models of DNA TEs, exemplified by *Tigger4*, *Tc1-10_Xt*, and *Tc1-2_ST*, respectively. Each dot shows the unmatched boundary where the non-autonomous copy is aligned to their consensus autonomous sequences. Dots with similar positions share similar boundaries. Each copy contains two unmatched positions on two sides. The copies are ordered by the left unmatched position.

(B) The overlaid structures of SB100X (light yellow) and *Tc1-2_ST* (sky blue) zooming on the SPDL motif (green) and DDE catalytic triad (red). Three novel residues, namely R131, R326, and G335 (marked in blue), are near the SPDL motif. Several hydrogen bonds (the dotted green line) are present in the middle. Additionally, W268 (also marked in blue) serves as a spatial neighbor to both the SPDL motif and D244.

(C and D) W195 (C) and H249 (D) are spatially close to DDE. Note that there is a hydrogen bond between H249 and D244. (C) also displays E154 (adjacent to D153).

(E) The overlaid structures of *hAT-7-PM* (pink), *HAT1_AG* (green), and *Hermes* (orange) zooming on the DDE catalytic triad (red). The figure convention follows (A) except that TIR sequences are represented by dark gray. C265, H268, and W319 (in RW motif; see also Figure 4E) are close to the DDE motif.

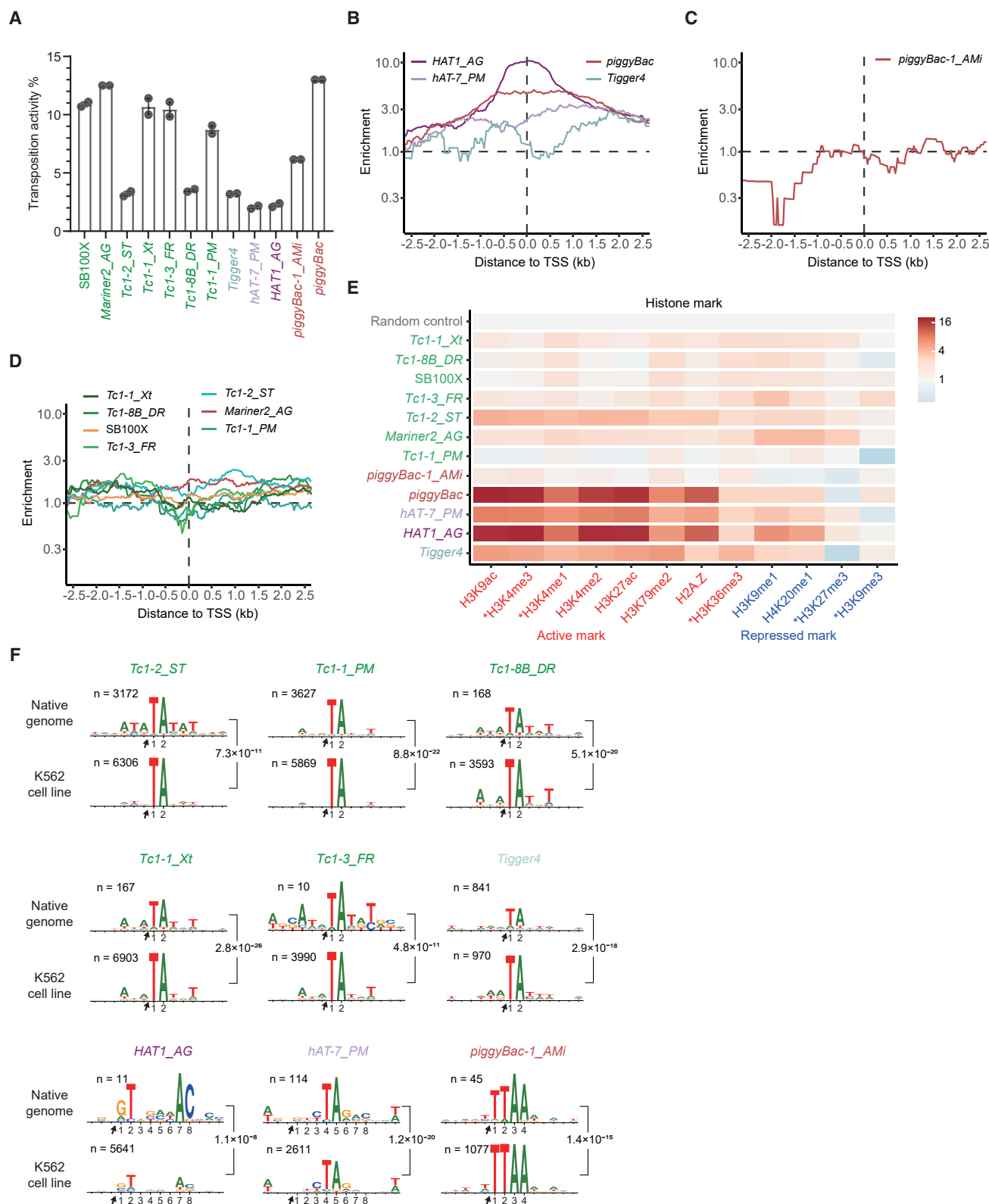
(F) The hydrogen bond between R318 and the third base, "C," on the transferring strands of TIRs, as well as the hydrogen bonds between S576 and surrounding residues (e.g., E572) or the third base "C" in TIRs. For R318, see also (E) and (H).

(G) R586 in the minor groove of TIRs.

(H) The hydrogen bond between S321 and R318.

(I) The length distribution of 40 active TEs across superfamilies or families (for *Tc1/mariner* and *hAT*).

The SB100X structure refers to PDB: 5cr4, whereas the structure of *Tc1-2_ST* was predicted by AlphaFold2. Analogously, the *Hermes* structure refers to PDB: 4d1q, whereas the structures of *hAT-7-PM* and *HAT1_AG* were predicted by AlphaFold2. The transposase-TIR complex models were generated using ChimeraX v1.4 (STAR Methods).



(legend on next page)

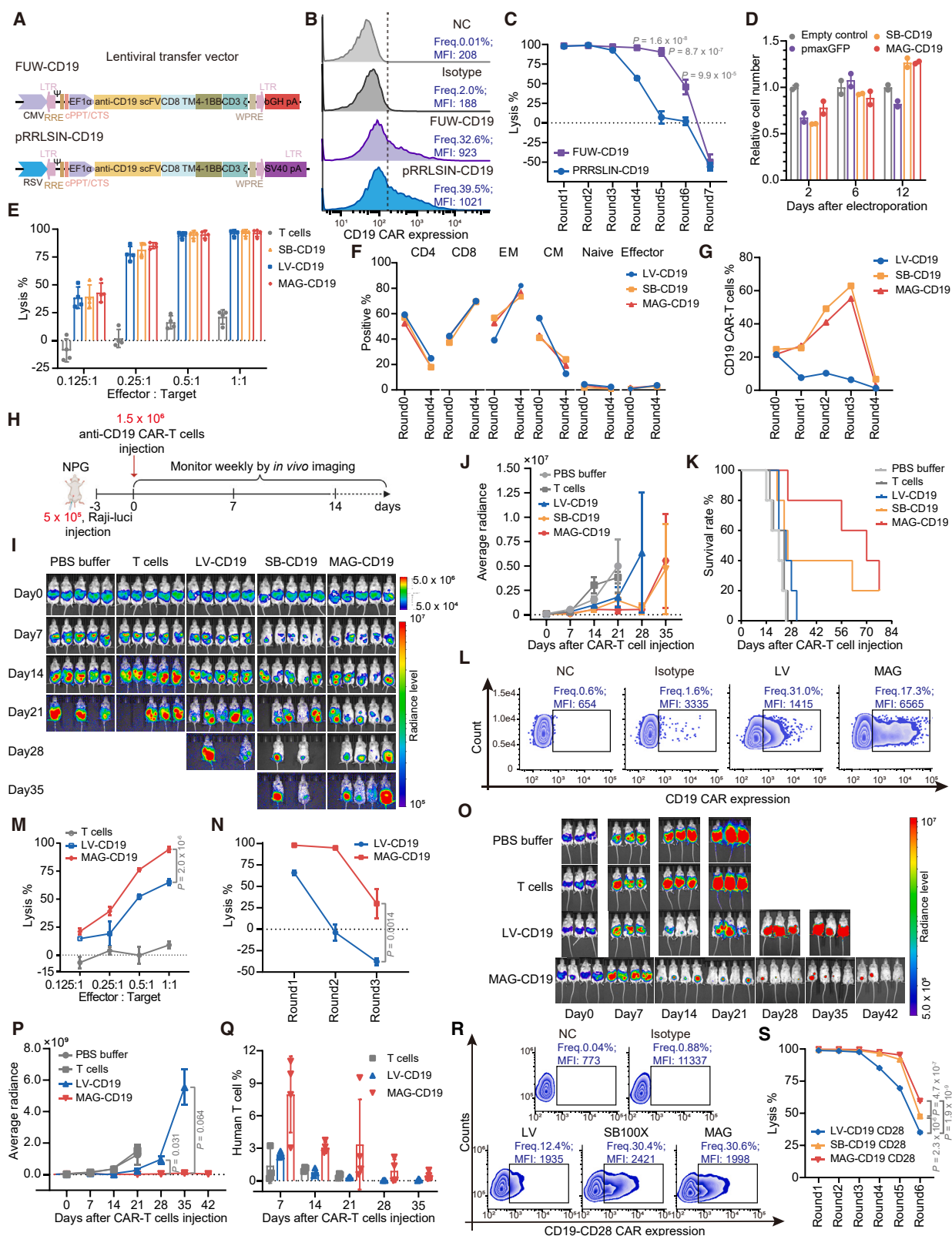
Figure S3. Functional characterization of representative active TEs, related to Figure 5

(A) The transposition activity of active TEs in K562 cells. The proportion of GFP-positive K562 cells 12 days after transfection represents the activity. Error bars represent the mean $\pm$ SEM of activities on the basis of two biological replicates, and the actual data points are overlaid.

(B–D) The fold enrichment of integration sites around TSSs in K562 cells. Four TEs including *piggyBac*, *HAT1_AG*, *hAT-7_PM*, and *Tigger4* show bias toward regions around TSSs (B), *piggyBac-1_Ami* avoids the upstream region of TSSs (C), and SB100X and the six *Tc1* TEs exhibit roughly random insertion patterns around TSSs (D).

(E) The integration fold enrichment heatmap of the top six active *Tc1* TEs, *Tigger4*, *HAT1_AG*, *hAT-7_PM*, *piggyBac-1_Ami*, SB100X, and *piggyBac* in K562 cells. Fold enrichment values were computed relative to histone marks, with TEs sorted as in Figure 5A. Within active and repressed mark groups, individual marks are ordered based on their decreasing mean enrichment fold. Note that ChromHMM utilized only five of the twelve marks, indicated with asterisks, to infer chromatin states (STAR Methods).⁷⁹

(F) The motifs at the integration sites in native and K562 genomes. The motifs are generated by WebLogo. Motifs from the two contexts are significantly similar (see also STAR Methods).



(legend on next page)

Figure S4. Anti-tumor capability of lentiviral and TE-derived CD19 CAR-T cells, related to Figure 6

(A) A schematic diagram illustrating two commonly used lentiviral vectors (FUW and pRRSLIN) for CD19 CAR. The CMV and Rous sarcoma virus (RSV) promoters were utilized, along with bGH and SV40 poly(A) signal sequences. TM denotes a transmembrane domain. ScFv, 4-1BB, and CD3 ζ represent three components of CAR: the single-chain variable fragment, the co-stimulatory domain, and the stimulatory domain, respectively. Functional elements of lentiviral vector are shown, including long terminal repeat (LTR), packaging signal of human immunodeficiency virus type 1 (Ψ), HIV rev response element (RRE), central polypurine tract and central termination sequence of HIV-1 (cPPT/CTS), and woodchuck hepatitis virus posttranscriptional regulatory element (WPRE).

(B) Flow cytometry analysis of CD19 CAR-T cells on the 12th day post-electroporation.

(C) Tumor cell lysis ability of CD19 CAR-T cells following multiple rounds of incubation with K562-CD19 cell lines. Data are represented as mean $\pm$ SEM ($n = 3$).

(D) The relative live T cell number. This value is defined as cell number in the case group divided by cell number of the empty control group (electroporation without any plasmid) on days 2, 6, and 12 after electroporation.

(E) Tumor cell lysis ability of CD19 CAR-T cells at different effector-to-target ratios ($n = 4$). The lysis ability is defined as follows: lysis % = $100 - ([RLU \text{ from wells with effector and target cell coculture}] / [RLU \text{ from wells with only target cells}] \times 100)$. Therefore, it is possible for the lysis percentage to be negative.

(F) T cell subtype composition of CAR-T cells before and after four rounds of incubation with K562-CD19 cell lines.

(G) The proportion of CD19 CAR-positive cells after one, two, three, and four rounds of incubation with K562-CD19 cell line. The initial proportion before incubation is also shown.

(H) Schematic of the *in vivo* experimental design using CDX models.

(I and O) Bioluminescent imaging results of NPG mice inoculated with Raji-luciferase tumor cells (Raji-luci) at different time points after PBS buffer, T cells, or CAR-T cells administration. (I) $n = 5$; (O) The group treated with MAG-CD19 CAR-T cells: $n = 4$; other groups: $n = 3$.

(J and P) Quantification of luminescence in (I) and (O).

(K) The survival curve of tumor-bearing mice treated with PBS buffer, T cells, or CAR-T cells.

(L) Flow cytometry analysis of CD19 CAR-T cells on the 12th day after electroporation.

(M) Tumor cell lysis ability of CD19 CAR-T cells with different effector-to-target ratios. Data are represented as mean $\pm$ SEM ($n = 4$).

(N) Tumor cell lysis ability of CD19 CAR-T cells after one, two, and three rounds of incubation with K562-CD19 cell lines. Data are represented as mean $\pm$ SEM ($n = 3$).

(Q) The proportion of human T cells in the peripheral blood of mice at different time points after T cells or CAR-T cells administration.

(R) CAR expression quantification by flow cytometry analysis of CD19-CD28 CAR-T cells on the 12th day following electroporation. Note that the CD28 co-stimulatory domain was employed rather than the 4-1BB domain.

(S) Tumor cell lysis ability of CD19-CD28 CAR-T cells following multiple rounds of incubation with K562-CD19 cell lines. Data are represented as mean $\pm$ SEM ($n = 3$).

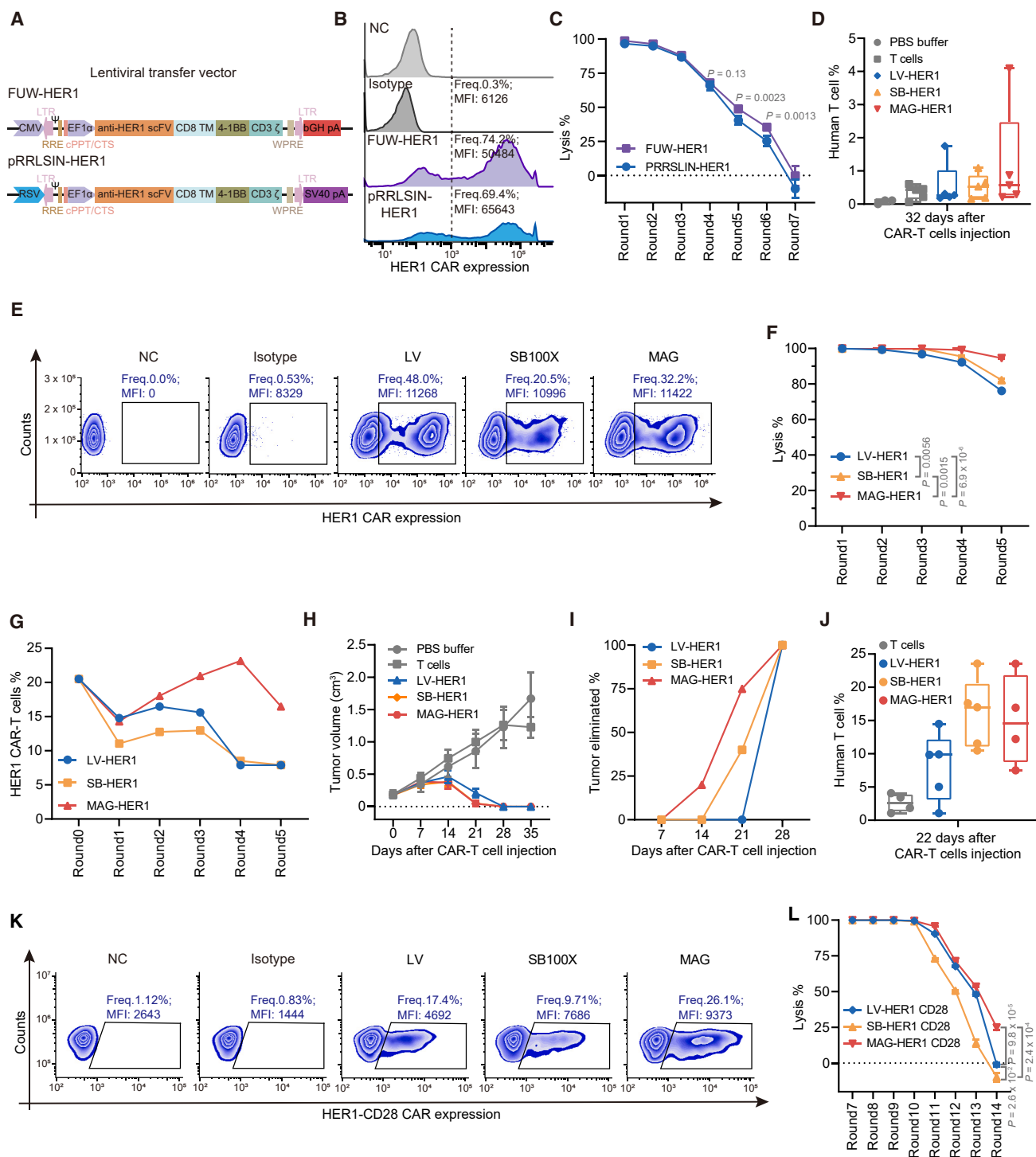


Figure S5. Anti-tumor capability of lentiviral and TE-derived HER1 CAR-T cells, related to Figure 6

(A) A schematic diagram illustrating FUW and pRRSLIN lentiviral vectors for HER1 CAR. The panel conventions follow Figures 6B and S4A. Notably, FUW-HER1 CAR-T cells are utilized in (D), (K), and (L), whereas pRRSLIN-HER1 CAR-T cells are employed in (E)–(J).

(B) CAR expression quantification by flow cytometry analysis of HER1 CAR-T cells on the 12th day post-electroporation.

(C) Tumor cell lysis ability of HER1 CAR-T cells following multiple rounds of incubation with H266 cell lines. Data are represented as mean $\pm$ SEM ($n = 3$).

(D) The proportion of human T cells in the peripheral blood of the CDX tumor model on the 22nd day following HER1 CAR-T cells injection. Data are represented as mean $\pm$ SEM ($n = 5$).

(E) CAR expression quantification by flow cytometry analysis of HER1 CAR-T cells on the 12th day post-electroporation.

(legend continued on next page)

-
- (F) Tumor cell lysis ability of HER1 CAR-T cells across 5 rounds of incubation with H266 cell lines. Data are represented as mean $\pm$ SEM ($n = 3$).
- (G) The proportion of HER1 CAR-T cells after one to five rounds of incubation with H266 cell line. The initial proportion is also shown.
- (H) Tumor volume of mice bearing HER1-positive tumors treated with PBS buffer, T cells, or CAR-T cells. Data are represented as mean $\pm$ SEM ($n = 5$).
- (I) The percentage of tumor-eliminated mice at 7, 14, 21, and 28 days following CAR-T cells injection.
- (J) The proportion of human T cells in the peripheral blood of the CDX tumor model on the 22nd day following HER1 CAR-T cells injection. Data are represented as mean $\pm$ SEM ($n = 5$).
- (K) Flow cytometry analysis of HER1-CD28 CAR-T cells on the 12th day post-electroporation. Note that the CD28 co-stimulatory domain was employed rather than the 4-1BB domain.
- (L) Tumor cell lysis ability of HER1-CD28 CAR-T cells following multiple rounds of incubation with H266 cell lines. Data are represented as mean $\pm$ SEM ($n = 5$).

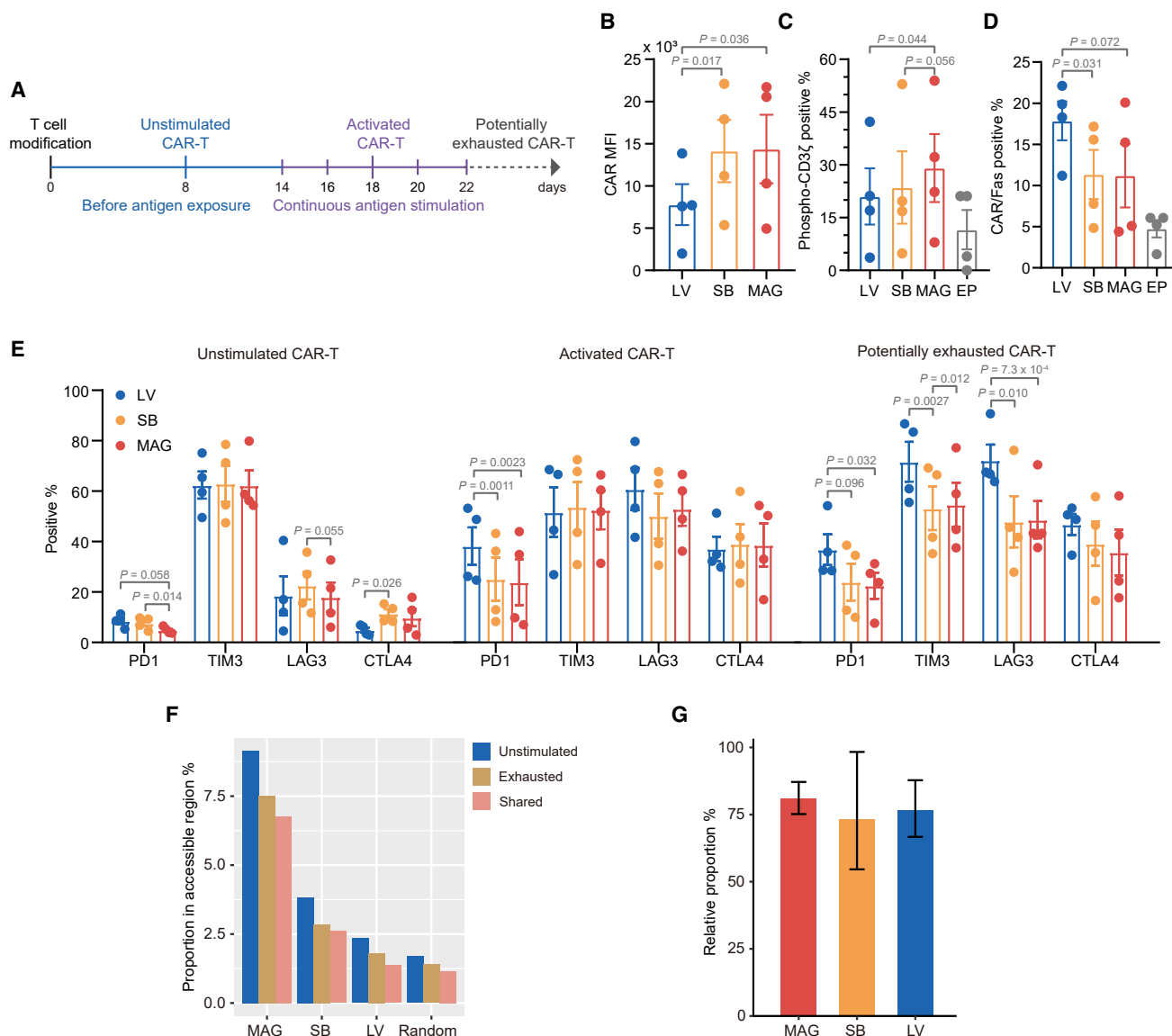


Figure S6. Potential mechanisms underlying the superiority of TE-based CAR-T cells, related to Figure 6

(A) Workflow for exhausting CAR-T cells. Unstimulated CAR-T cells were cultured for 14 days before exposure to tumor cells or antigen. Activated and potentially exhausted CAR-T cells were generated by stimulation through two rounds of fresh tumor cell challenges and at least four rounds of tumor cell challenges, respectively.

(B) CAR expression quantification. Flow cytometry analysis of CAR expression in unstimulated CAR-T cells on the 8th day post-electroporation ($n = 4$, with two replicates for anti-CD19 cells and two replicates for anti-HER1 cells). Data are represented as mean $\pm$ SEM.

(C) Proportion of Phospho-CD3 ζ and CAR double-positive cells. Phospho-CD3 ζ reflects the extent of tonic signaling. "EP" indicates empty control (electroporation without any plasmid). Data are represented as mean $\pm$ SEM ($n = 4$).

(D) Proportion of Fas and CAR double-positive cells. Fas reflects the extent of AICD. Data are represented as mean $\pm$ SEM ($n = 4$).

(E) Proportion of exhaustion marker-positive cells in unstimulated, activated, and potentially exhausted CAR-T cells. Four markers including PD1, TIM3, LAG3, and CTLA4 were used. Data are represented as mean $\pm$ SEM ($n = 4$).

(F) Proportion of integration sites in open chromatin regions of unstimulated CAR-T cells and exhausted CAR-T cells. "Shared" represents integration sites located within open chromatin regions in both types of CAR-T cells. Error bar represents the 95% CI. Accessible chromatin regions were identified based on assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq) data (see also STAR Methods).

(G) The proportion of integration sites within open chromatin regions in exhausted CAR-T cells over those in unstimulated CAR-T cells. Error bars represent the 95% confidence interval (CI).

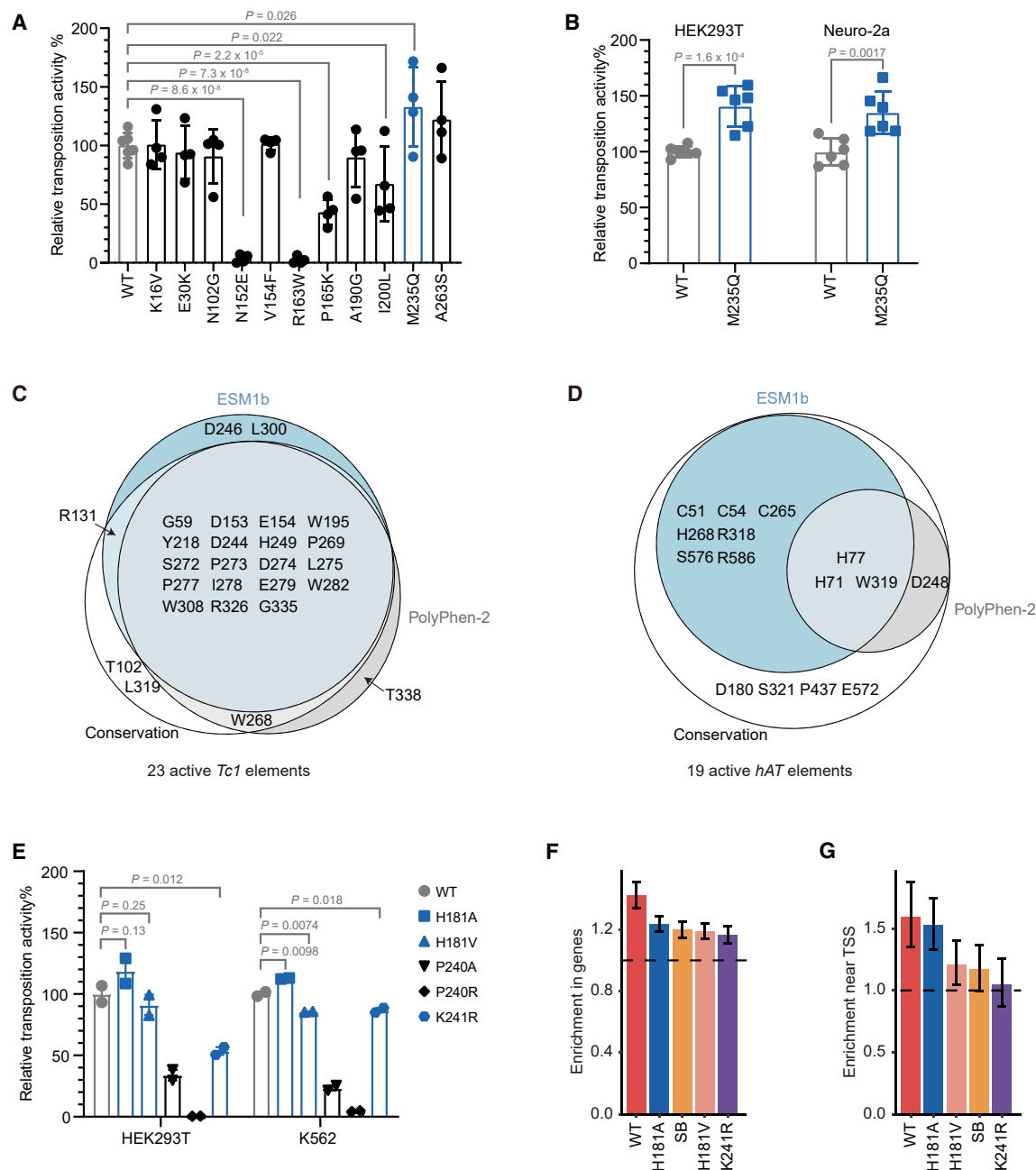


Figure S7. A pilot attempt to optimize MAG and assess the performance of tools in predicting crucial residues, related to Figures 4 and 5

(A) The relative transposition activity of 11 MAG mutations in HEK193T cell line. Among 46 sites shared across more than 50% *Tc1* elements (Figure 4A) but changed in MAG, we randomly picked 11 sites and mutated them to the conserved state. Each error bar represents the mean $\pm$ SD of relative transposition activity on the basis of four biological replicates. The actual data points are also overlaid. Only p values smaller than 0.05 are shown.

(B) The relative transposition activity of M235Q mutant in human HEK193T and mouse Neuro-2a cell lines. For M235Q, we performed validation experiments in both HEK193T and Neuro-2a cell lines. Each error bar represents the mean $\pm$ SD of relative transposition activity on the basis of six biological replicates. The actual data points are also overlaid.

(C and D) Venn diagram illustrating critical residues identified in 23 active *Tc1* elements (C) or 19 active *hAT* elements (D) by the conservation-based method, PolyPhen-2, and ESM1b. "Conservation" denotes residues invariant across all active TEs.

(E) Relative transposition activity of wild-type MAG and five mutants in HEK293T and K562 cells. Data are represented as mean $\pm$ SD ($n = 2$).

(F and G) Fold enrichment of integration sites for wild-type MAG, SB, and the three MAG mutants in gene regions (F) and TSS proximal (1 kb) regions (G) in K562 cells. Dashed lines represent random expectations (1.0), and error bars represent the 95% CI. TEs are sorted by decreasing folds. Since P240A and P240R mutations in (E) heavily reduced the activity of MAG in both cell lines, they were excluded from insertion site mapping analyses.