Science Advances
AAAS

Supplementary Materials for

Programmable RNA-guided DNA endonucleases are widespread in eukaryotes and their viruses

Kaiyi Jiang *et al.*

Corresponding author: Omar O. Abudayyeh, omar@abudayyeh.science; Jonathan S. Gootenberg, jgoot@mit.edu

Sci. Adv. **9**, eadk0171 (2023)
DOI: 10.1126/sciadv.adk0171

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

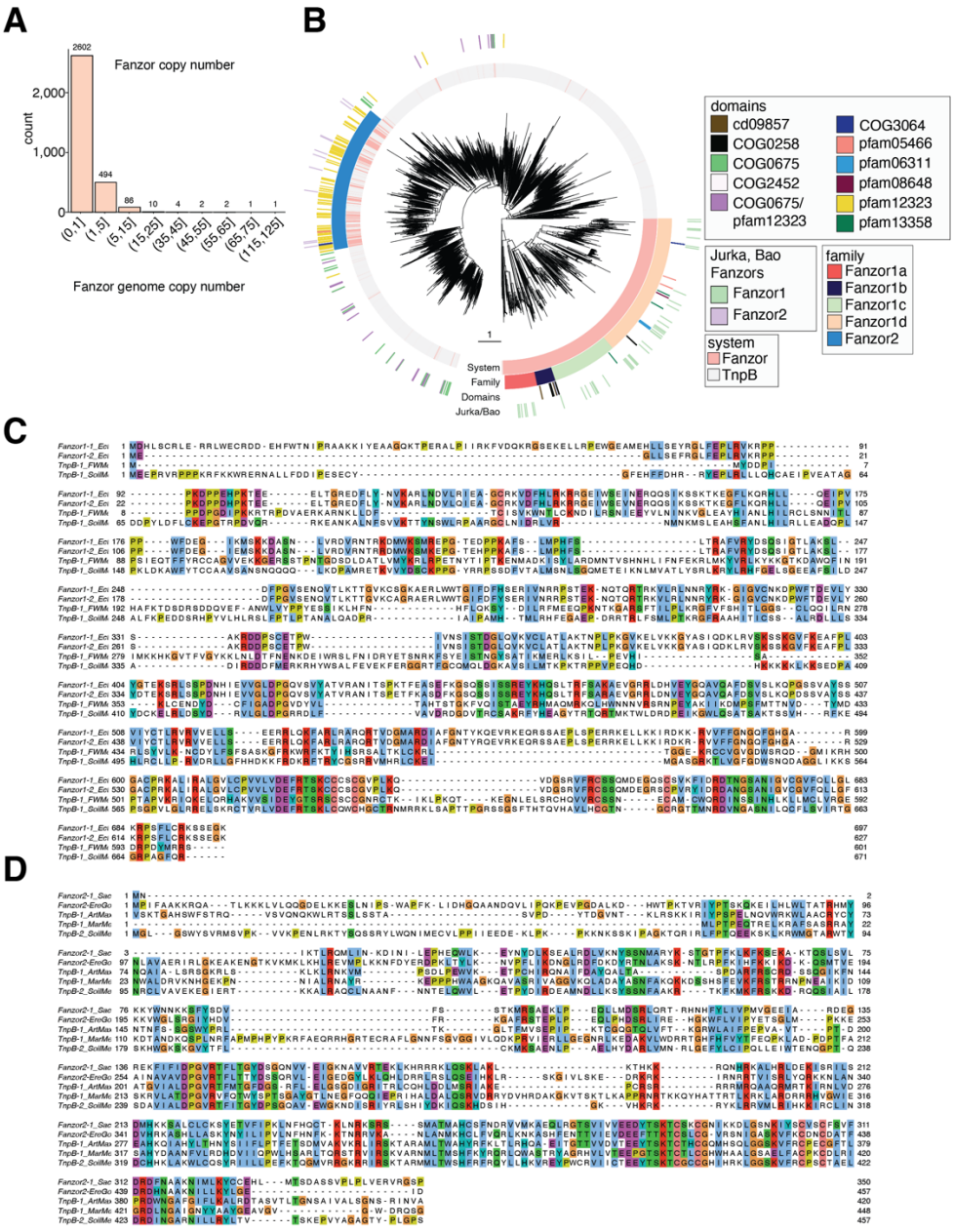

Fig. S1: Genomic characteristics of Fanzor family members

A) A histogram of the copy number of individual Fanzor members inside their respective genomes. B) Phylogenetic tree of Fanzors and TnpBs with the domain predictions of nearby proteins marked as a ring (the nearest 5 genes downstream and upstream). Previously discovered Fanzors are marked in the outer ring*(5)*. C) Alignment of Fanzor1 proteins with closely related TnpBs. D) Alignment of Fanzor 2 proteins with closely related TnpBs.

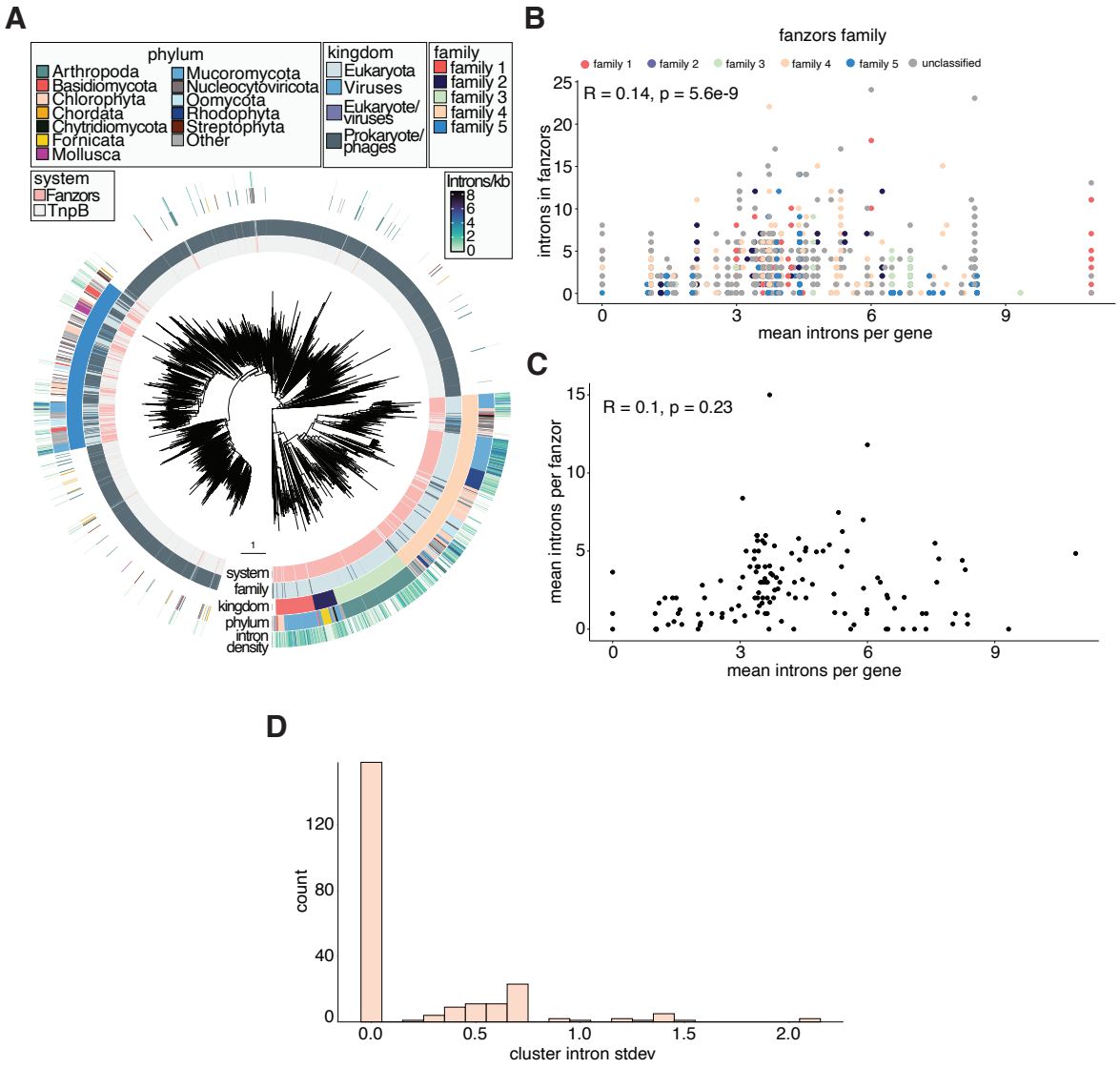

Fig. S2: Fanzor intron characterization.
A) Phylogenetic tree of Fanzors and TnpBs with rings to show the host superkingdom, phylum, and intron density of the Fanzor proteins. B) Scatterplot showing the intron density of the Fanzor proteins along with the mean intron density of their host genomes. Fanzor proteins are colored according to their family designations. C) Scatterplot showing the mean intron densities of the Fanzor proteins in a genome along with the mean intron density of their host genomes. D) Histogram of the standard deviation of intron densities within 70% similarity clusters of Fanzor proteins.

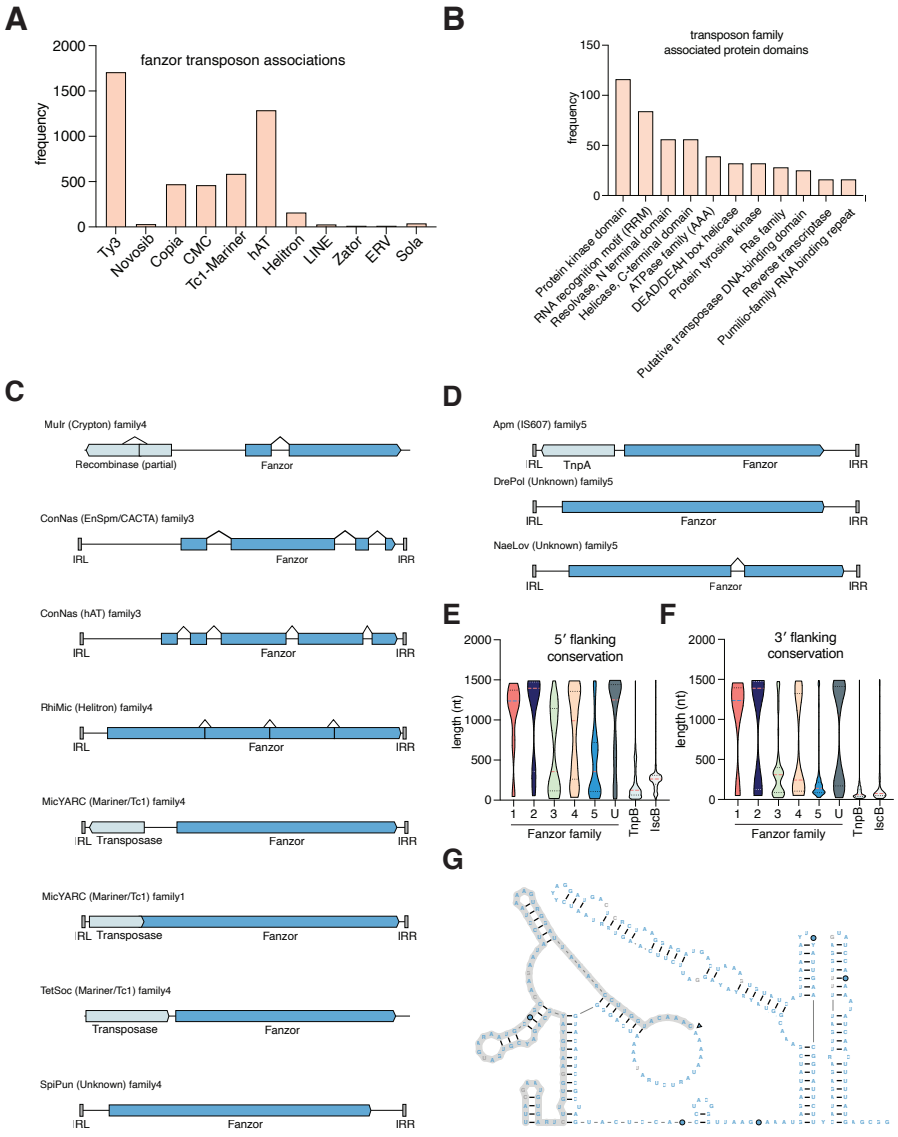

Fig. S3: Locus characteristics of Fanzor family members
A) Frequency of predicted associated transposons nearby Fanzor (within +/– 10 kb) per transposon family type. B) Frequency of the top occurring nearby protein domains within 5 genes upstream or downstream of the Fanzor MGE. C) Locus schematics of different Fanzor1 nucleases and their associated transposons. IRL marks the left inverted repeat and IRR marks the right inverted repeat. D) Locus schematics of different Fanzor2 nucleases and their associated transposons. E) Comparison of predicted flanking non-coding conservation lengths at the 5′ end of the MGEs of IscB, TnpB, and each Fanzor family. F) Comparison of predicted flanking non-coding conservation lengths at the 3′ end of the MGEs of IscB, TnpB, and each Fanzor family. G) Conserved secondary structure of fRNAs between the different copies of the ApmFNuc family. Shaded gray area corresponds to conserved sequence not present in the mature fRNA, potentially removed by RNase processing (cut site designated by blue triangle).

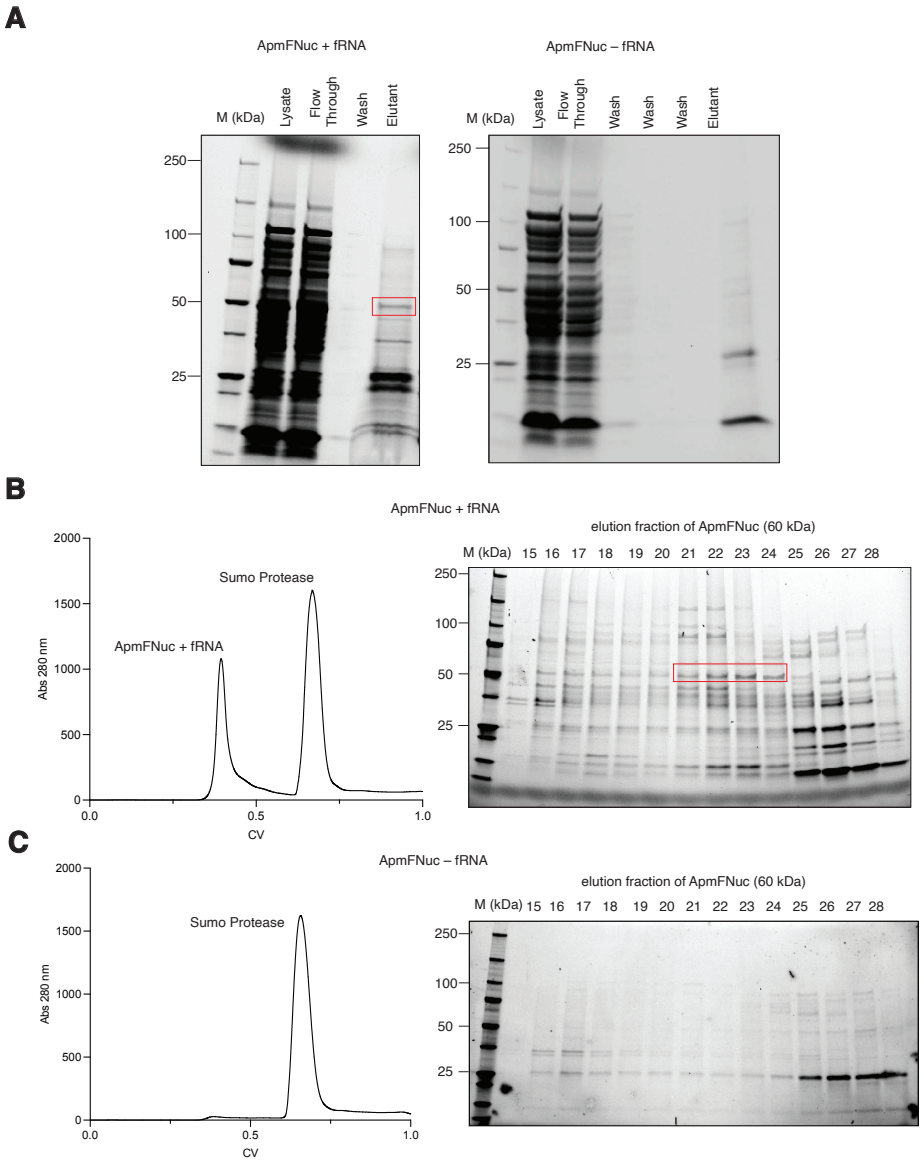

Fig. S4: Purification of ApmFNuc RNPs.
A) Protein gel showing flowthrough and eluent of ApmFNuc products during gravity flow strep-bead purifications prior to loading of FPLC. Red square denotes the desired protein product. B) FPLC traces of ApmFNuc purified with its fRNA and protein gels showing each fraction's protein products with the desired protein product that was pooled labeled with red squares. C) FPLC traces of ApmFNuc purified without its fRNA and protein gels showing no RNP product in all observed fractions.

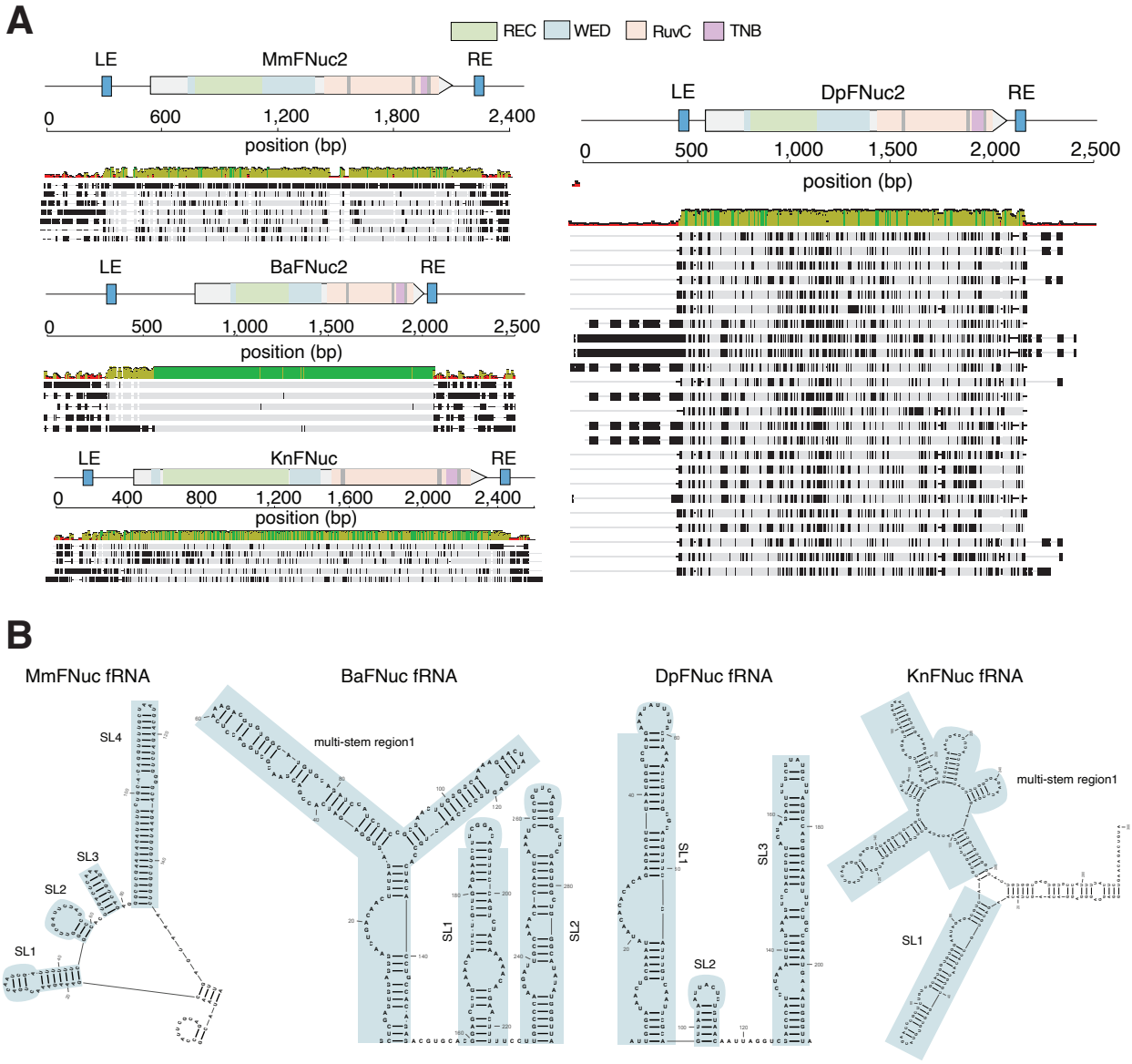

Fig. S5: Characterization of eukaryotic Fanzor nucleases.
A) Alignment and domain annotation of three eukaryotic Fanzor nucleases (DpFNuc, MmFNuc, and BaFNuc). RE and LE elements are determined by conservation dropoff between alignments of different copies in the genome. B) Secondary structure prediction of fRNAs associated with DpFNuc, MmFNuc, and BaFNuc determined by small RNA sequencing of the locus. Blue shaded region denotes stem loops and multi-stem loops region in the fRNAs.

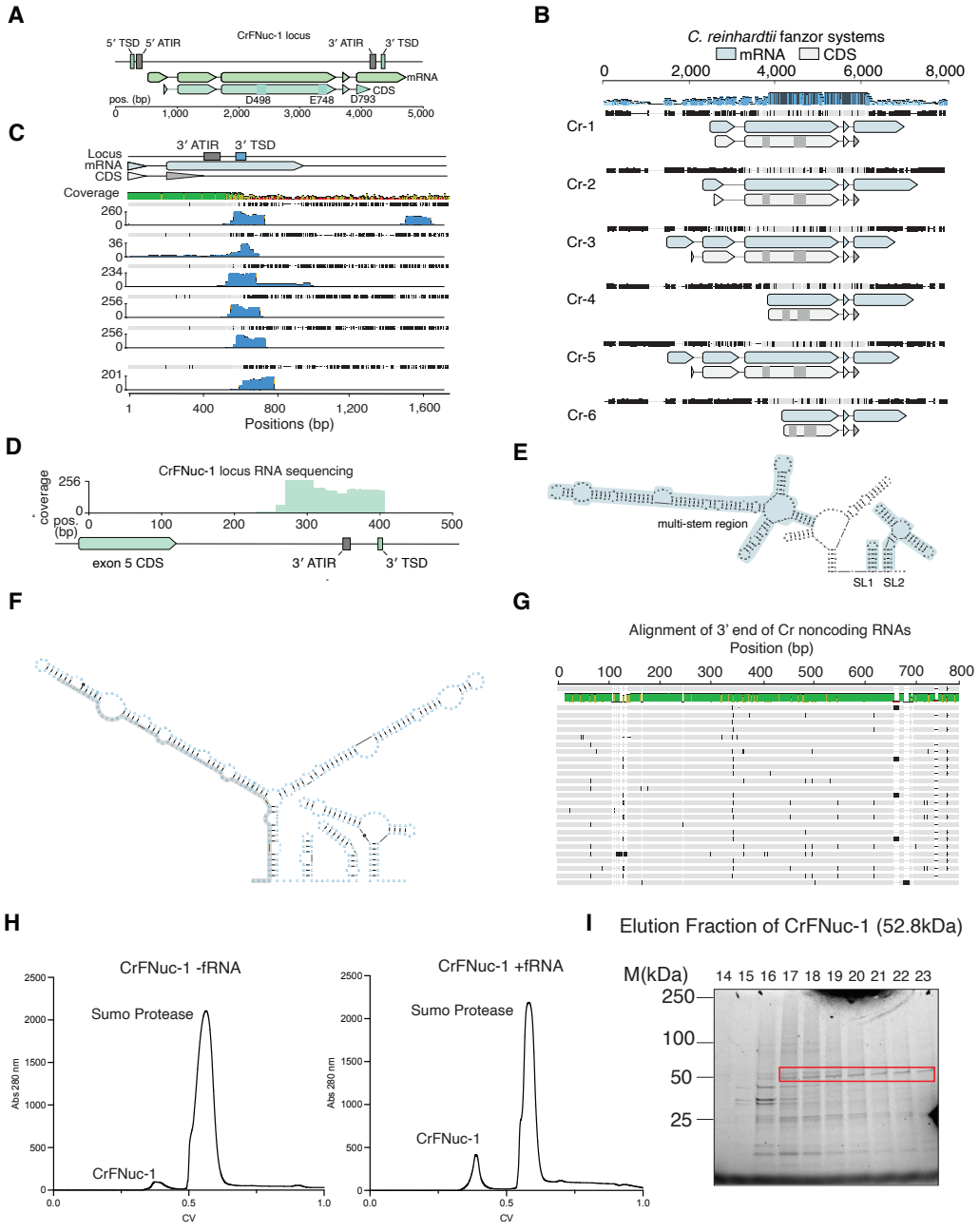

Fig. S6: Characterization of Cr-1FNuc and its fRNA.
A) Schematic of the *Chlamydomonas reinhardtii* Fanzor1 system (Cr-1FNuc), including the 5′ asymmetrical terminal inverted repeats (ATIR), 3′ ATIR, 5′ target site duplications (TSD), 3′ TSD, and the mRNA and coding sequences for Cr-1FNuc. The mRNA track shows the processed mRNA transcripts relative to the genome and the CDS track shows the ORF coding sequences relative to the genome. B) Alignment of all six copies of Fanzor systems inside the annotated parts of *C. reinhardtii* genome, showing highly conserved 3′ ends of the CrFNuc proteins along with their fRNAs and variable 5′ end compositions of the proteins. The blue track shows the processed mRNA transcripts relative to the genome and the gray track shows the ORF coding

sequences relative to the genome. C) Small RNA sequencing traces mapped onto all 6 copies of RuvC-containing Fanzor systems in the *C. reinhardtii* genome. D) Small RNA sequencing of the *Chlamydomonas reinhardtii* organism showing expression of a noncoding RNA species at the 3′ end of the Cr-1FNuc locus that extends beyond the ATIR into the TSD. E) Secondary structure of Cr-1FNuc non-coding RNA from Fig. 3J, showing significant folding of the fRNA. F) Conserved secondary structure of the six CrFNuc fRNA copies in the genome. G) Alignment of the 26 full or partial copies of Fanzor MGEs inside the *C. reinhardtii* genome at their 3′ ends. H) FPLC traces of Cr-1FNuc purified either with or without its fRNA, showing that the RNP complex is only stable when the correct fRNA is expressed and present. The Cr-1FNuc peak in the FPLC trace is labeled. I) Protein gel showing elution fractions of the Cr-1FNuc with the desired protein product that was pooled labeled with a red square.

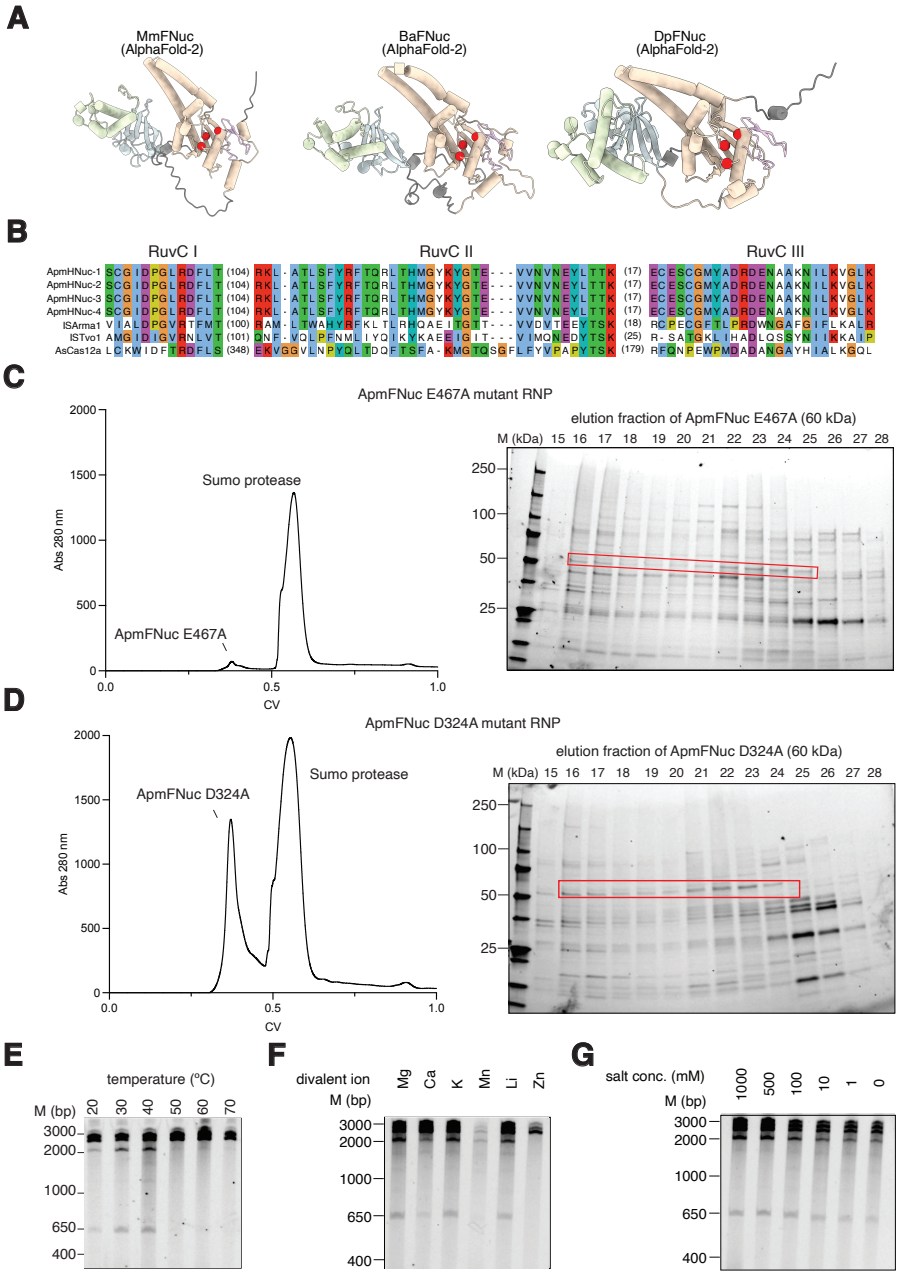

Fig. S7: Further characterization of ApmFNuc nuclease activity.
A) Predicted AlphaFold-2 structures of MmFNuc, DpFNuc, and BaFNuc showing that despite having a rearranged glutamate in the RuvC catalytic domain, the catalytic aspartates and glutamates form a putative active catalytic triad (red residues). B) Alignment of ApmFNuc Ruvc domain with Isdra2TnpB RuvC domain to nominate the catalytic RuvC-I aspartic acid (D324) and the RuvC-II glutamic acid (E467A). C) FPLC traces of ApmFNuc E467A mutant purified with its fRNA and protein gels showing each fraction's protein products with the desired protein product that was pooled shown with a red square. D) FPLC traces of ApmFNuc D324A mutant purified with its fRNA and protein gels showing each fraction's protein products with the desired protein product that was pooled shown with a red square. E) Native TBE gel showing nuclease activity of ApmFNuc at temperatures from 10 to 65 degrees Celsius. Reactions were carried out by incubating wild-type ApmFNuc RNP on a plasmid with the TGGG TAM 5′ adjacent to the 21

nt spacer target. Cleavage was visualized by gel electrophoresis. F) Native TBE gel showing nuclease activity of ApmFNuc with different cations supplemented into the cleavage buffer. Reactions were carried out by incubating wild-type ApmFNuc RNP on a plasmid with the TGGG TAM 5′ adjacent to the 21 nt spacer target. Cleavage was visualized by gel electrophoresis. G) Native TBE gel showing nuclease activity of ApmFNuc with different NaCl salt concentrations supplemented into the cleavage reaction buffer. Reactions were carried out by incubating wild-type ApmFNuc RNP on a plasmid with the TGGG TAM 5′ adjacent to the 21 nt spacer target. Cleavage was visualized by gel electrophoresis.

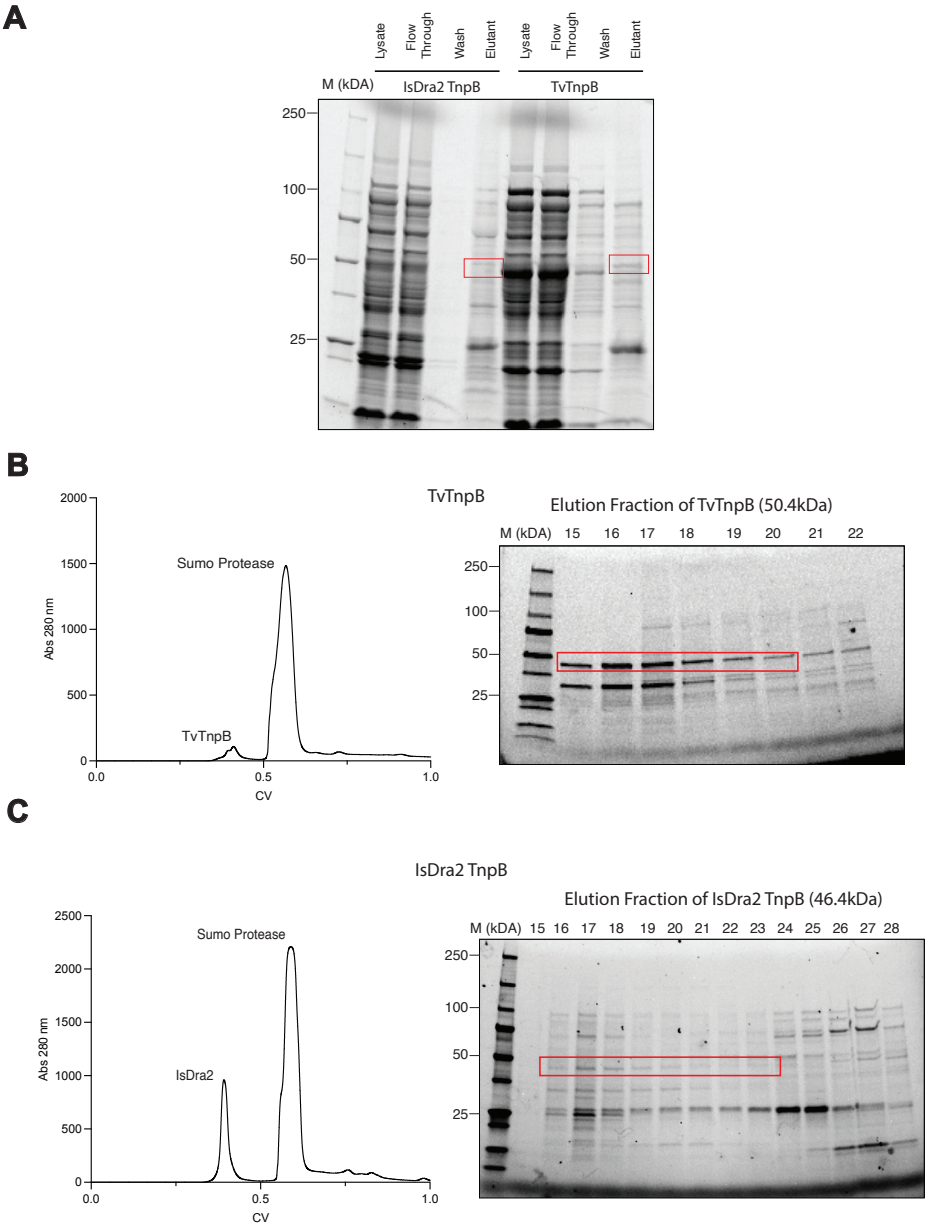

Fig. S8: Purification of Isdra2TnpB and TvTnpB.
A) Protein gel showing flowthrough and eluent fractions of Isdra2TnpB and TvTnpB products during gravity flow strep-bead purifications. The desired protein product is shown via a red square. B) FPLC traces of TvTnpB purified with its ωRNA and protein gels showing each fraction's protein products with the desired protein product that was pooled shown with a red square. C) FPLC traces of Isdra2TnpB purified without its ωRNA and protein gels showing each fraction's protein products with the desired protein product that was pooled shown with a red square.

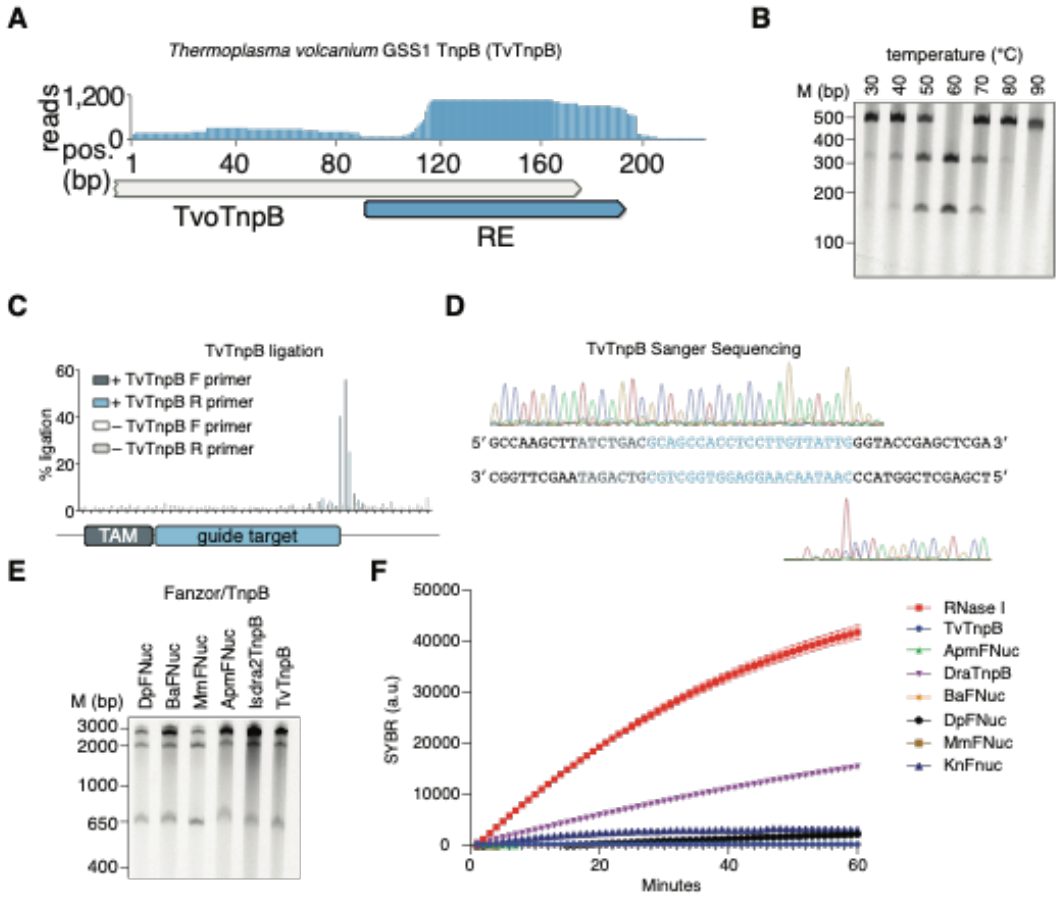

Fig. S9: Characterization of TvTnpB and collateral activity comparisons.
A) Expression of the non-coding RNA for TvTnpB, revealing a specific non-coding RNA species that associates with the TvTnpB protein extending from the ORF to outside the RE element similar to Isdra2TnpB. B) TvTnpB DNA cleavage of a 21 nt target containing a 5′ ATGAC TAM at temperatures ranging from 30 degrees Celsius to 90 degrees Celsius, showing optimal cleavage reaction temperature near 60 degrees for TvTnpB. C) Next-generation sequencing mapping of the TAM cleavage by TvTnpB via adaptor ligation. Reads were aligned to the TAM target to map cleavage locations. Two separate reactions were ran in parallel with and without addition of TvTnpB RNP. The cleavage products were amplified in both 5′ and 3′ directions with F denoting 3′ direction and R denoting the 5′ direction. D) Sanger sequencing traces of TvTnpB cleavage on a 5′ CTGAC TAM target, showing cleavage at the end of the target. E) On target cleavage activity of TvTnpB, Isdra2TnpB, MmFNuc, BaFNuc, DpFNuc, and ApmFNuc. Nucleases were incubated with plasmids containing their preferred TAM site and on-target guide RNA sequences for 1 hour of cleavage and subsequently visualized on a native TBE gel for comparison of on-target cleavage activity. F) Fluorescent signal from RNase alert reporter detection of RNA collateral cleavage activity from RNase A, TvTnpB, Isdra2TnpB, MmFNuc, BaFNuc, DpFNuc, and ApmFNuc incubated with their target DNA sequences for 1 hour. The signal is normalized to a no DNA target condition

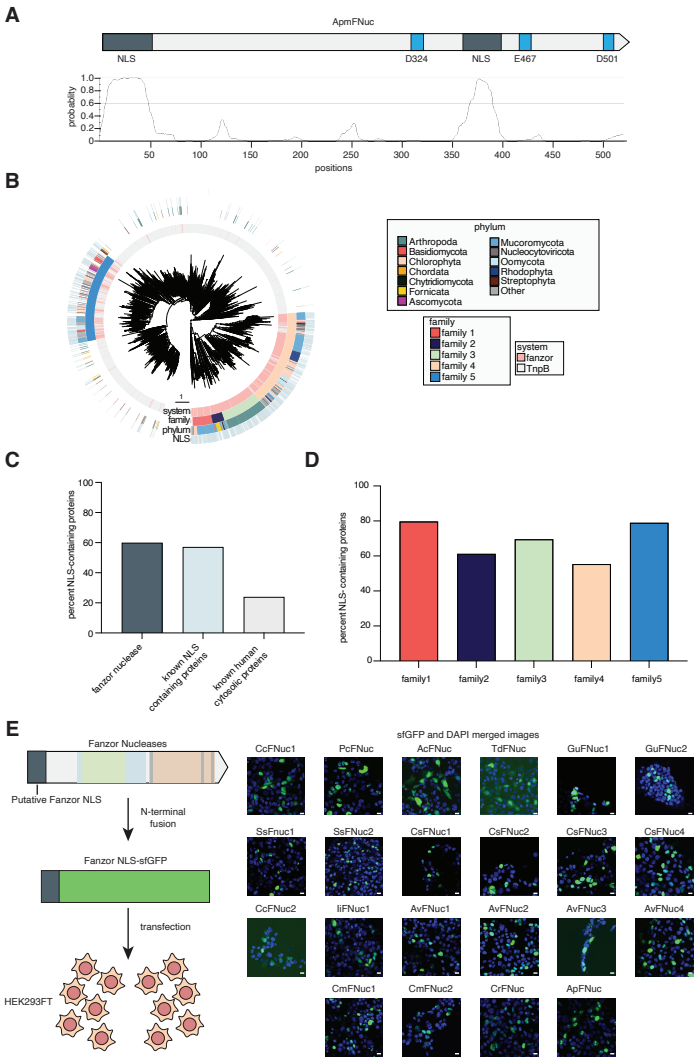

Fig. S10: Characterization of Fanzor nuclear localization signals.
A) Probability distribution of potential NLS elements across the ApmFNuc protein sequence as predicted by NLStradamus*(13)*. The default cutoff at 0.6 is used to call significant NLS like elements, revealing one N-terminal NLS and one internal NLS. B) Phylogenetic tree of Fanzor nucleases and TnpB orthologs, with rings marking the host phyla and family designations of the Fanzor orthologs and which proteins were predicted to have an NLS sequence.C) A bar plot depicting NLS predictions rates on a set of known human cytosolic proteins (negative control), a set of known NLS containing proteins (positive control), and all Fanzor nucleases. D) Per family breakdown of NLS containing Fanzor predictions for Fanzor families 1-5. E) Confocal images of 22 different Fanzor nuclease N-terminal NLS predictions fused to sfGFP and transfected into HEK293FT cells for visualization of nuclear localization of the sfGFP. DAPI is used to stain the nucleus and images are shown with the GFP and DAPI channel signals merged. Scale bar, 20 μm.

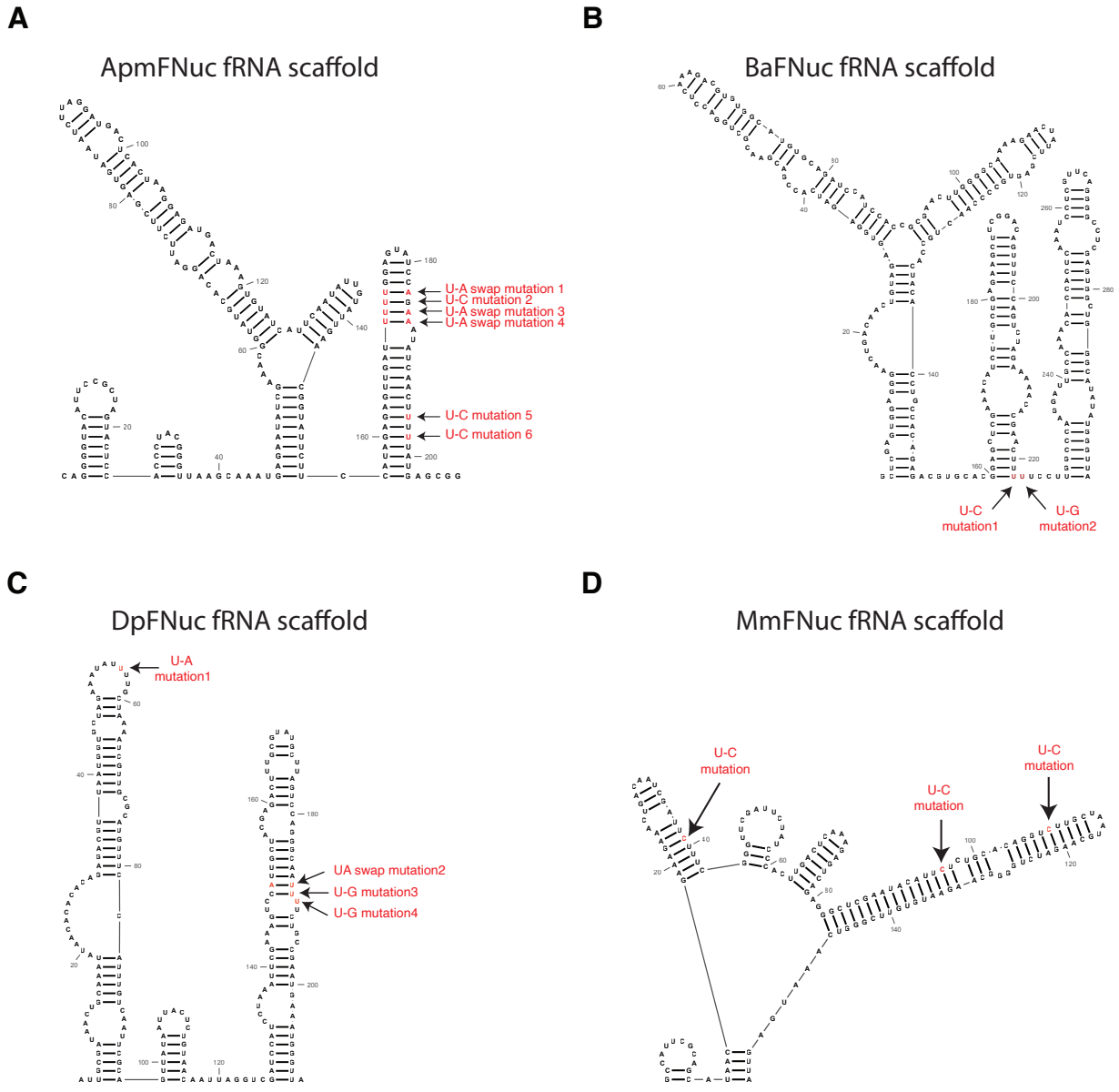

Fig. S11: Schematic of engineered fRNA scaffolds for mammalian genome editing. fRNA secondary structures are predicted by viennaRNA fold for A) ApmFNuc, B) BaFNuc, C) DpFNuc and D) MmFNuc. Mutated residues are labeled in red color and the arrows pointing to each base denote the nucleic acid mutations introduced at the specific position.

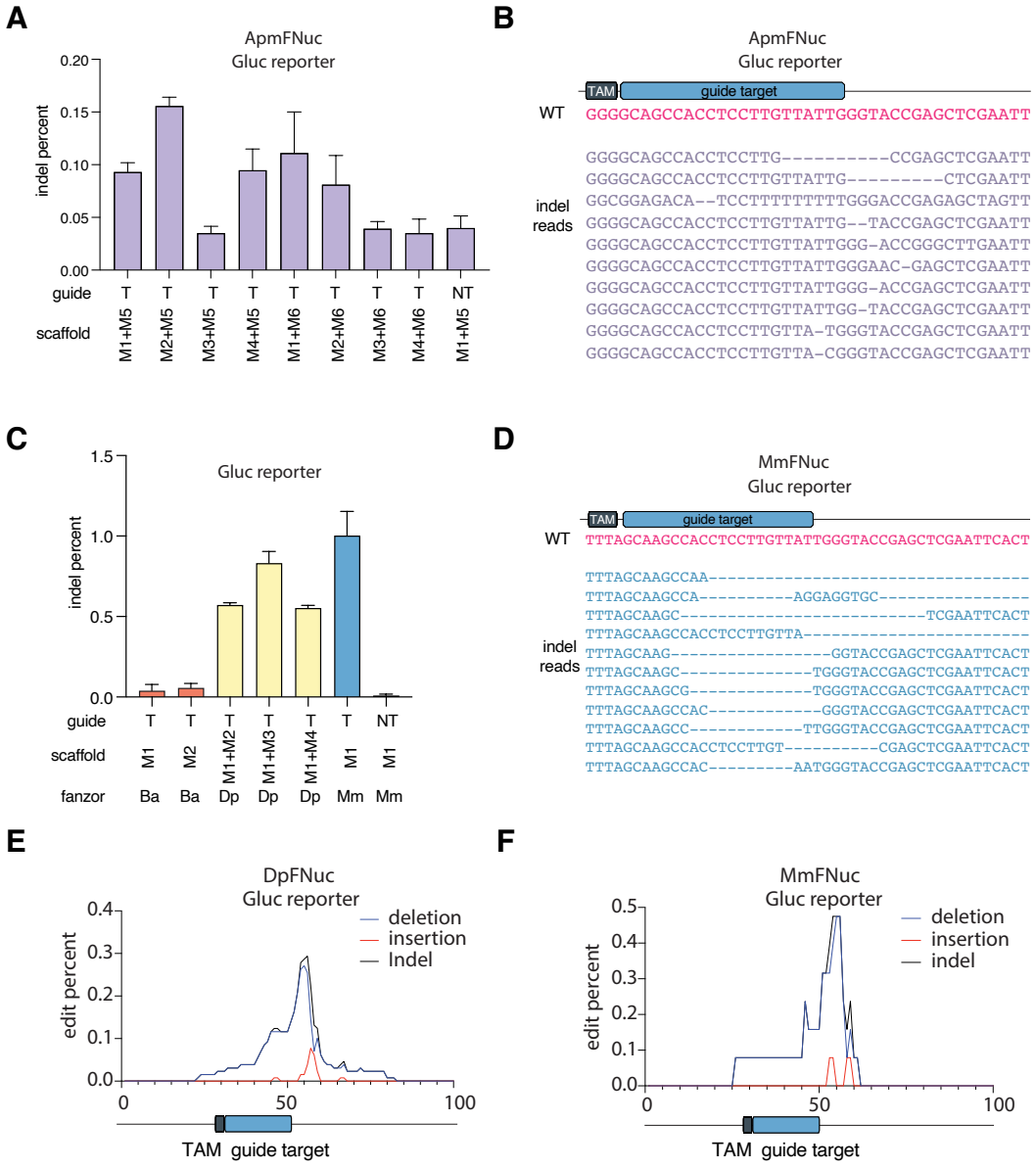

Fig. S12: Characterization of Fanzor nuclease plasmid reporter editing in HEK293FT cells. A) An ApmFNuc mammalian expression vector and its fRNA U6 expression plasmid are co-transfected into HEK293FT cells targeting a luciferase plasmid reporter. Different mutations on the wild-type fRNA scaffold are introduced as shown in fig. S11 to eliminate poly-U stretches in the fRNA. Indel frequency is measured by next-generation sequencing with targeted primers on the plasmid reporter. B) Representative indel alleles from the M2+M5 scaffold targeting guide condition on the luciferase reporter, showing deletions centered around the 3′ end of the guide target. C) Indel frequency on the luciferase plasmid reporter for BaFNuc, MmFNuc, and DpFNuc with different engineered fRNA scaffolds. D) Representative indel alleles for MmFNuc with the M1 fRNA scaffold targeting the luciferase reporter plasmid, showing deletions centered

around the 3′ end of the guide target. E) Quantification of insertion, deletion and combined indel frequencies generated on the plasmid reporter by DpFNuc with the (M1+M3) scaffold targeting guide condition. Rates are shown per base throughout the quantification window of the amplicon. F) Quantification of insertion, deletion and combined indel frequencies generated on the plasmid reporter by MmFNuc with the targeting guide condition. Rates are shown per base throughout the quantification window of the amplicon.

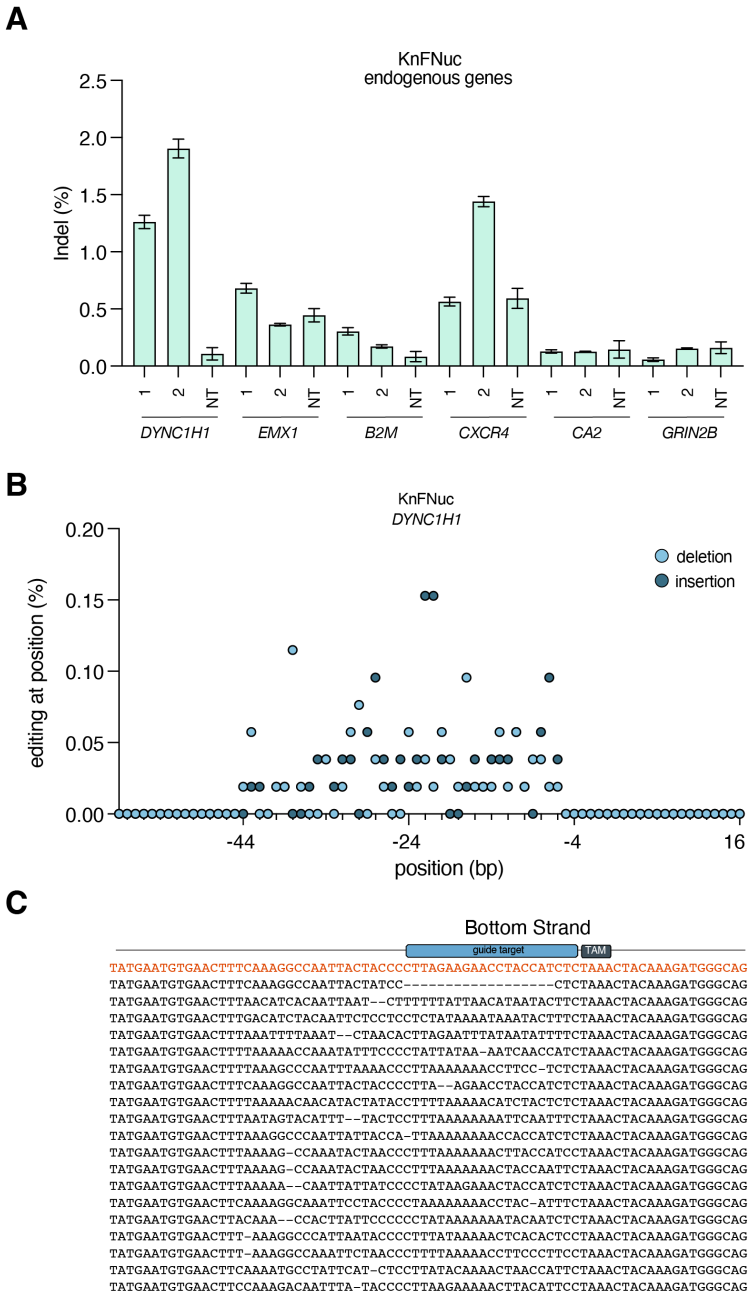

Fig. S13: Characterization of KnFNuc Fanzor1 nuclease genomic editing in HEK293FT cells.

A) A KnFNuc mammalian expression vector and its fRNA U6 expression plasmid are co-transfected into HEK293FT cells targeting 6 different genomic targets. Indel frequency is measured by next-generation sequencing with targeted primers on the target. B) Quantification of insertion and deletion frequencies generated on the *DYNC1H1* genomic target by KnFNuc. Rates are shown per base throughout the quantification window of the amplicon. C) Representative indel alleles showing deletions and insertions centered around the 3′ end of the guide target.

Table S1. *Fanzor* families in eukaryotics genomes and their identified transposon associations.
Fanzor elements are named after the host species. *Fanzor2* elements are indicated by *. The left and right termini are indicated by L. and R. respectively, in the orientation of the encoded *Fanzor* protein. N: none; n.a.: not available; i.c.: incomplete. [#]: The encoded Tpase (or coding sequences). If a given *Fanzor* element does not encode Tpase, but the superfamily it belongs can be determined, the superfamily name is parenthesized. Rows highlighted in white correspond to Fanzor-Transposon associations previously identified *(5)*. Rows highlighted in orange correspond to new transposon associations identified in this study.

Family (bp)	Copy No.	Termini	TIR (bp)	TSD (bp)	Fanzor protein (aa) &(No. Exons)	Tpase [#] (Superfamily)	Comments
MDe-1	2				815 (3)		
MDe-2	2				698 (4)		
MDe-3	1				620 (4)		
MDe-4	4	L.R.	N	n.a.	731 (4)		
MDe-5	4	L.R.	N	n.a.	656 (4)		
MDe-6 (3852)	10	L.R.	N	n.a.	661 (4)		
MDe-7 (3937)	8	L.R.	24	2 (TA)	772 (3)		
MDe-8	4	R.			745 (3)		
MDe-9	3	R.			764 (5)		
MDe-10	1				779 (3)		
MDe-11	3	R.			713 (4)		
MDe-12 (3875)	5	L.R.	N	n.a.	677 (4)		
MDe-13	3	R.			680 (2)		
HMa-1	1				i.c.	Mariner	Probably from virus
*SAl-1**	3	R.			400 (1)		
*SAl-2**	3	R.			498 (4)		
SPu-1 (2149)	25	L.R.	33	2 (TA)	633 (1)		
SPu-2	2				663 (1)		
SPu-3 (2288)	2	L.R.	25	2 (TA)	626 (1)		
ROr-1 (5190)	10	L.R.	90	2 (TA)	928 (3)	Mariner	

ROr-2 (4073)	18	L.R.	46	2 (TA)	690 (2)	Mariner	
ROr-3 (2862)	16	L.R.	133	2 (TA)	720 (2)		
ROr-4 (5244)	9	L.R.	38	9	1165 (3)	(MuDr)	
AMa-1	1				871 (4)		
AMa-2	1				645 (3)		
AMa-3	1				789 (7)		
PBl-1 (3938)	4	L.R.	12	3 (TAN)	683 (4)		
PBl-2	3				677 (2)		
PBl-3 (4614)	6	L.R.	42	9	1186 (3)	(MuDr)	
MCi-1 (4036)	4	L.R.	20	2 (TA)	686 (2)	Mariner	
MCi-2A (10235)	3	L.R.	N	11	1375 (4)	Crypton	
MCi-2B	2	R.	-	-	1375 (4)		
MCi-2C	3	R.	-	-	1375 (4)		
MCi-2D (9295)	2	L.R.	N	12	1375 (4)		
MCi-3 (5305)	2	L.R.	39	4? (TTAA)	1304 (2)		
MCi-4 (4508)	6	L.R.	31	9	1245 (3)	(MuDr)	
MCi-5 (7323)	5	L.R.	N	n.a.	1212 (3)	Harbinger	
MCi-6	2				1231 (2)		
MCi-7	1	R.			1153 (3)		
MCi-8	1				1067 (2)		

MCi-9	1				1149 (3)		
MCi-10	1				1135 (4)		
*AGo-1**	1				457 (1)		
*ECy-1**	1				455 (1)		
*SCe-1**	1				350 (1)		
*TDe-1** (1785)	7	L.R.			486 (1)		
DFa-1 (11949)	12	L.R.	12	4	1241 (10)	(Sola2)	
DFa-2 (12887)	7	L.R.	12	4	1010 (9)	Sola2	
DFa-3 (10254)	2	L.R.	13	4	1084 (10)	(Sola2)	
DFa-4	1				1020 (13)	-	
PPa-1 (13566)	3	L.R.	22	4	1699 (7)	Sola2	
PPa-2	1				945 (8)		
PPa-3	1				970 (9)		
PPa-4 (14423)	3	L.R.	16	4	1827 (14)	Sola2	
PPa-5 (15292)	3	L.R.	16	4	1388 (12)	Sola2	
PPa-6	2	R.	16	4	1218 (13)		
PPa-7	1				1756 (16)		
*ACa-1** (2675)	2	L.R.	N	0	603 (1)	TnpA_IS607	
*ACa-2**	1				653 (1)	TnpA_IS607	
VCa-1	1				768 (1)		
VCa-2	1				i.c.		
CRe-1 (3992)	>100	L.R.	N	0 or n	830 (5)	(Helitron)	Expressed

CRe-2 (4882)	>100	L.R.	N	0 or n	906 (10)	(Helitron)	Expressed
CRe-3 (4688)	>100	L.R.	N	0 or n	967 (10)	(Helitron)	Expressed
CRe-4	3	R.			944 (6)		
CRe-5	3	R.			i.c.		
CVu-1	n.a				i.c.		
CMe-1A (3169)	150	L.R.	N	n.a.	734 (1)		
PUl-1 (3620)	8	L.R.	24	2 (TA)	802 (1)	Mariner	
PUl-2 (3820)	1	L.R.	33	2 (TA)	643 (3)	Mariner	
PUl-3	1				799 (1)		
PUl-4 (3356)	3	L.R.	26	2 (TA)	809 (1)		
PUl-5	1	R.			617 (1)		
PUl-6	5	R.			642 (1)		
NOc-1	4				i.c.		
PSo-1	2	R.			660 (1)		
PSo-2	4	R			726 (1)		
PSo-3	3				716 (1)		
PSo-4	3				785		
*PSo-5**	1				i.c.		
PCa-1,	2	R.			788 (1)		
PCa-2 (2107)	2	L.R.	N	N	611 (1)		
*PCa-3**	2	R.			483		
PRa-1	1				i.c.		
*PRa-2**	2	R.			i.c.		
ALa-1	1				i.c.		
ALa-2	1				i.c.		
ESvi-1A (3180)	1	L.R.	59		890 (1)		
ESvi-1B (4052)	1	L.R.	25	8	890 (1)	IS4	
ESv-1 (2639)	2	L.R.	40	2 (TA)	693 (1)		
ESv-2 (3603)	2	L.R.	18		757 (1)	IS4	
SWv-1 (2633)	1	L.R.	21	6	779 (1)		

HAgv-1 (1963)	2	L.R.	13	4 (TTAT)	572 (1)		
HAmv-1 (1925)	1	L.R.	13	4 (TTAA)	592 (1)		
PUgv-1 (1961)	2	L.R.	13	4 (TTAT)	571 (1)		
SFav-1 (1954)	2	L.R.	13	4 (TTAN)	606 (1)		
HVav-1 (1955)	5	L.R.	13	4 (TTAN)	608 (1)		
MCnv-1	1	R.			i.c.		
PGv-1 (4442)	1	L.R.	29	2 (TA)	625 (1)	Mariner	
EHv88-1	1				650 (1)		
*EHv99B1-1** (2126)	1	L.R.			640 (1)		
*ISvMimi_1** (2549)	3	L.R.			520 (1)	TnpA_IS607	*=APmv-2, =ACmv-2*
*ISvMimi_2**	1				545 (1)	TnpA_IS607	*=APmv-1, =ACmv-1*
*APmv-3**	1				482 (1)		*=ACmv-3*
MGvc-1,*	1				526 (1)		
*MGvc-2**	1				493 (1)		
*ISvAR158_1**	1				351 (1)	TnpA_IS607	
*ISvNY2A_1** (2164)	3	L.R.			395 (1)	TnpA_IS607	
*ISvNY2A_2** (1443)	2	L.R.			432 (1)		
*CRv-1**	1				416 (1)	TnpA_IS607	
*FEsv-1**	1				408 (1)		
Fanzor1-1_SitMos	>12	L.R.	11-bp	2-bp (NN)	(3)	EnSpm?	
Fanzor1-2_SitMos	>8	L.R.	74	8 (ATGTANNN)	(5)	hAT	
Fanzor1-3_SitMos	>14	L.R.	12	8	(5)	hAT	
Fanzor1-4_SitMos	1	-	-				fragmental
Fanzor1-5_SitMos	6	R.	-	-		Helitron?	
Fanzor1-6_SitMos	>10	L.R.	21	2-bp (NN)		EnSpm?	
Fanzor1-7_SitMos	6	L.R.	127	8 GT(GTGNNNNN)	(4)	hAT	
Fanzor1-8_SitMos	>7	L.R.	12	2-bp (NN)	(3)	EnSpm?	
Fanzor1-9_SitMos	>16	R.	-	-	(4)		fragmental
Fanzor1-10_SitMos	>9	R.	-	-			fragmental
Fanzor1-11_SitMos	1	-	-				
Fanzor1-1_ConNas	>20	L.R.	12	2		EnSpm?	
Fanzor1-2_ConNas	>6	L.R.	12	2		EnSpm?	

Fanzor1-3_ConNas	>50	L.R.	12	2	(2)	EnSpm?	
Fanzor1-4_ConNas	>20	L.R.	11	2	(3)	EnSpm?	
Fanzor1-5_ConNas	>7	L.R.	11	2	(3)	EnSpm?	
Fanzor1-6_ConNas	>10	L.R.	133	8 (ATGTANNN)	(5)	hAT	
Fanzor1-7_ConNas	>3	L.R.	126	8 (GTGNNNNN)	(3)	hAT	
Fanzor1-8_ConNas	>8	L.R.	12	2	(4)	EnSpm?	
Fanzor1-9_ConNas	>13	L.R.	126	8 (ATGTANNN)	(3)	hAT	
Fanzor1-10_ConNas	>6	L.R.	none	8 (GCANNNNN)	(4)	hAT?	
Fanzor1-11_ConNas	>10	L.R.	133	8 (ATGTANNN)	(5)	hAT	
Fanzor1-12_ConNas	>10	L.R.	130	8 (ATGTANNN)	(3)	hAT	
Fanzor1-13_ConNas	>11	L.R.	72	2 (TA)	(4)	EnSpm?	
Fanzor1-14_ConNas	>4	L.R.	12	2 (TA)	(3)	EnSpm?	
Fanzor1-15_ConNas	>3	L.R.	16	8 (GGTANNNN)	(1)	hAT?	
Fanzor1-16_ConNas	>3	L.R.	none	8 (GGTANNNN)	(6)	hAT?	
Fanzor1-17_ConNas	>2	L.R.	15	8 (GGTANNNN)	(3)	hAT?	
Fanzor1-18_ConNas	>20	R.					
Fanzor1-19_ConNas	>4	L.R.	121	8 (ATGTANNN)	(4)	hAT	
Fanzor1-1_ApoVar	>16	L.R.	none	0	(7)	Crypton	
Fanzor1-2_ApoVar	12	L.R.	none	0		Crypton?	
Fanzor1-3_ApoVar	>4	L.R.	none			Helitron	
Fanzor1-4_ApoVar	>11	L.R.	none			Crypton	
Fanzor1-5_ApoVar	>6	L.R.	none			Helitron	
Fanzor1-6_ApoVar	>6	L?.R.	none			Helitron?	

Fanzor1-7_ApoVar	>5	L.R.	none			Helitron	
Fanzor1-8_ApoVar	>5	L.R.		TA		Mariner	
Fanzor1-8B_ApoVar	>5	L.R.		TA		Mariner	
Fanzor1-9_ApoVar (3996)	=4	L.R.	19-bp	TA		Mariner?	
Fanzor1-1_RhiMic	3	L.R.	90	2 (TA)	(1)	Mariner?	
Fanzor1-2_RhiMic	>3	L.R.	none	2 (TA)	3	Mariner (+)	
Fanzor1-3_RhiMic	>4	L.R.	none			Helitron	
Fanzor1-4_RhiMic	~4	L.?R?	none				
Fanzor1-1_MuIr	~3	R.		0		Crypton	
Fanzor1-2_MuIr	~4	L.R.	36	9	(5)	MuDR?	
Fanzor1-3_MuIr	>4	R.					
Fanzor1-4_MuIr	>3	L.R.		9		MuDR?	
Fanzor1-5_MuIr	>4	L.R.	Weak subterminal TIRs	9		MuDR?	
Fanzor1-1_ParPar	>10	L.R.	none			Crypton	
Fanzor1-2_ParPar	>10	L.R.	142	2 (TA)			
Fanzor1-3_ParPar	>3	L.R.	24	3(TWA)			
Fanzor2-1_ParPar (1660)	>40	L.R.	14	4 (TTAA)			
Fanzor1-1_KleNit	>6	L.R.	27	2 (TA)	(1)	Mariner	
Fanzor1-1_KleNit	>5	L.R.	27	?			
Fanzor1-1_ChlPri	>4	L.?R?					
Fanzor2-1_ChlPri (2654)	>23	L.R.	13	5			
Fanzor1-1_CarMem	=6	L.R.	none	5	1		
Fanzor1-2_CarMem	>6	L.R.	none	5	1		

Fanzor1-3_CarMem	=3	L.R.		5			
Fanzor1-1_MicYARC (3453)	>100	L.R.	27	2(TA)	(1)	Mariner (+)	Target CTA
Fanzor1-1N1_MicYARC	>14	L.R.	27	2(TA)	(1)	Mariner	
Fanzor1-2_MicYARC		L.R.	27	2(TA)	(1)	Mariner	Target CATA
Fanzor1-3_MicYARC	>16	L.R.		2(TA)		Mariner	
Fanzor1-4_MicYARC	>50	L.R.	32	2(TA)		Mariner (+ strand)	Target GTTA, specific
Fanzor1-5_MicYARC	>2	L.R.		2(TA)	(1)	Mariner(- strand)	Target CATA, specific
IS607EU-1_MicYARC	>20	L.R.	none	none		IS607, S-recombinase	
IS607EU-1_MicYARC (2163)		L.R.				IS607, S-recombinase	
Fanzor1-1_XesXan	>4	L.R.		TTAA		piggyBac (by TIR)	
Fanzor1-1_CycCry	>9	L.R?	none		>3		88%
Fanzor1-1_EreLig	=3	L.R.	17	4-bp	1	piggyBac?	
Fanzor1-1_AbrTri (1873)	=7	L.R.	13-bp	4-bp		?	
Fanzor1-1_CydSpl (1931)	=5	L.R.	12-bp	4-bp	1		
Fanzor1-1_NeHa	>6	L.R.		none	1	Crypton??	14642-bp
Fanzor1-2_NeHa	>3	L.R.		none			
Fanzor1-1_HypPro	>4	L.R.	9	TTAA	1	piggyBac ?	Inserted with I-element.
Fanzor1-1_LysCor (2202)	=3	L.R.	10	TTAA	1	piggyBac ?	
Fanzor1-1_NeYa	>40	R.					

IS607EU-h1_PhySoj	>2	L.? R.					
Fanzor1-6_PhySoj (2476)	>2	R.					
IS607EU-1_UndPin	>3(*)	indeterminate				IS607	Integrated inside MuDR
Fanzor1-1_LepBou	>3	L.R.	24-bp	TA	1	Mariner (by TIR)	Target TGTA
Fanzor1-2_LepBou	=2	L.R.	33-bp	Mostly TA	1	EnSpm (by TIR)	
Fanzor1-3_LepBou	=2	L.R.		2-bp	1	EnSpm (by TIR)	
IS607EU-1_GiMa						IS607	
IS607EU-2_GiMa	>60	L.R.	none	none		IS607	TnpB degraded.
IS607EU-3_GiMa	>14	L.R.	none	none		IS607	
Fanzor1-1_PilApi	>40	L.R.	18-bp	4-bp			
Fanzor1-2_PilApi	=8	L.R.	none	none			
Fanzor1-3_PilApi	>6	L.R.	169-bp	TA, likely			Old repeat, 86% identity
IS607EU-1_SchTIO01	>20	L.R.	none	none		IS607	
Fanzor1-1_VerVer	>28	L.R.	20	2-bp			
Fanzor1-1_EuLap	>7	R.					
Fanzor1-1_GuiThe (2751)	=9	L.R.	15-bp	4-bp (ATAN)			
Fanzor1-2_GuiThe (2714)	>10	L.R.	none	4-bp (TTAW)			TnpB truncated at the C-terminal
Fanzor1-3_GuiThe (2261)	=1	L.R.	18-bp	4-bp (predicted)			
Fanzor1-1_ApoBC	~4	R.	uncertain	uncertain			
Fanzor1-1_AphGif	=8	R.	uncertain	Uncertain			5'-end is flexible.
Fanzor1-1_HemFu	=9	L.R.	14-bp	4-bp TTAA		piggyBac (by TIR)	

Fanzor1-1_MucSat	>9	L.R.	27-bp	2-bp (TA)			
Fanzor1-1_BomMaj	>4	R.	uncertain	uncertain			
Fanzor1-2_BomMaj	=3	R.	uncertain	uncertain			
Fanzor1-1_RhiDel	=3	L.R.	78-bp	TA?		Mariner?	TGTA
Fanzor2-1_MerMer	=4	R.				IS607?	

Table S2. Fanzor Protein and fRNA sequences used in this study

Fanzor/TnpB systems	Fanzor/TnpB types	Protein Sequence	Associated fRNA Scaffold Sequence for biochmiestry (neglecting guide)
ApmFNuc	Fanzor2	MKEAVKNVKPKVPAKKRIITGSK TKKKVFVKKKPPDKKPLKKPVKKT VKTYKLKSIYVSNKDLKMSKWIPT PKKEFTEIETNSWYEHRKFENPN GSPIQSYNKIVPVVPPESIKQQNL ANKRKKTNRPIVFISSEKIRIYPTKE QQKILQTWFRLFACMYNSSIDYI NSKKVVLESGRINVAATRKVCNKI SVRKALKTIRDNLIKSTNPSIMTHI MDEAIGLACSNYKTCLTNYIEGQI KKFDIKPWSISKRRKIIVIEPGYFK GNSFCPTVFPKMKSSKPLIMIDKT VTLQYDSDTRKYILFVPRVTPKYS VNKEKNSCGIDPGLRDFLTVYSE NETQSICPIEIVVNTTKNEYKKIDK INEIIKTKPNLNSKRKKKLNRGLRK YHRRVTNKMKDMHYKVSHELV NTFDKICIGKLNVKSILSKANTVLK SALKRKLATLSFYRFTQRLTHMGY KYGTEVVNVNEYLTTKTCSNCGKI KDLGASKIYECESCGMYADRDEN AAKNILKVGLKPWYKQK	AAAAATAGTCTAATAAAATCAGGGGTACA TTCCGCTAGTACTCCACCCTACGGGTTAAG CAAATGAGAATATCGAAACGGTATGCACA GGATTCTTCGAGTGATAATCTTAGGATGAC TCACTAAGGAGATGACTAAAGTGTATCATT CAATATTGTATTGAACGGTATTCTTCCATA GAGAGTTGATTTtTGGAGTATCCAGAAATA TCAACTtTTTATGAGCGG
DpFNuc	Fanzor2	MKRKREDLTLWDAANVHKHKS MWYWWEYIRRKDMVNHEKTD CDVIQLLQSASVKKQKTQSDKFLT SFSVGIRPTKHQKRVLNEMLRVS NYTYNWCLWLVNEKGLKPHQFE LQKIVCKTNANDVDPQYRMEND DWFFNNKMTSVKLTSCKNFCTS YKSAKSLKSKLKRPMSVSNIIQGS FCVPKLFIRHLSSKDVSTDNTNM QNRYICMMPDNFEKRSNPKERF LKLAKPITKIPPIDHDVKIVKRADG MFIMNIPCDPKYTRRNASNDTIE KRVCGIDPGGRTFATVYDPIDCC VFQVGIKEDKQYVISKLHNKIDHA HMHLTKAQNKKQQQAARERIVS	ATTGGATGTTCAAAATGAAGCATACACTTC GAAGACGTGTGGAGTGTGTGGAACAATA AACAAAAATCTAGAAAAGAGTGAAACATT TTATTGCGATAACTGCAAATATAACACACA CAGAGACGTTAATGGTGCTAGAAATATtTT GCTAAAATCGTTGCGCATGTTTCCATTTGT CAATTCGCAGTTATAATTACTCTGTAACAA TTAGGTCGATCCATCCTAAATTCGAAAGTC CATTGCTACGAGACTTTGCGTATGCTTAGT CCAGGGCAATtTTCTGCCGAATGAAATGG GTTA

		LKKTHLKLKTFVDDIHLKLSSHLVK EYQYVALGKINVAQLVKTDRPKP LSKRAKRDLLYWQHYRFRQRLTH RTTNTECILDVQNEAYTSKTCGV CGTINKNLEKSETFYCDQCKYNT HRDVNGARNILLKSLRMFPFEKQ QQ*	
MmFNuc	Fanzor2	MKRKREQMTLWKAAFVNGQET FKSWIDKARMLELNCDVSSASST HYSDLNLKTKCAKMEDKFMCTFS VGIRPTSKQKRTLNQMLKVSNH AYNWCNYLVKEKDFKPKQFDLQ RVVTKTNSTDVPAEYRLPGDDW FFDNKMSSIKLTACKNFCTMYKS AQTNQKKTKVDLRNKDIAMLRE GSFEVQKKYVRLLTEKDIPDERIR QSRIALMADNFSKSKKDWKERFL RLSKNVSKIPPLSHDMKVCKRPN GKFVLQIPCDPIYTRQIQVHTSDSI CSIDPGGRTFATCYDPSNIKAFQI GPEADKKEVIHKYHEKIDYVHRLL AYAQKKKQTQAVQDRIGQLKKL HLKLKTYVDDVHLKLCSYLVKNYK LVVLGKISVSSIVRKDRPNHLAKK ANRDLLCWQHYRFRQRLLHRVR GTDCEAIAQDERYTSKTCGNCGV KNNKLGGKETFICESCNYKTHRD VNGARNILCKYLGLFPFAA*	ACTTCCAAGACCTGTGGTAATTGCGGTGT GAAGAACAACAAACTTGGTGGAAAGGAA ACGTTTACTTGTGAGTGTTGCAATTACAAA ACTCATCGAGACGTCAACGGAGCGAGAAA CATTCTGTGCAAATACTTGAAACTTTTTCCA TTCGCAGCATAACGAAAGAAACTGACAAT CGATTTTTTCGGGTTCGATTCTATCCCACTT GACTCAAAGAGTCAGAGGGCTCGAATACA TTTTCTGCACAGGTTTTGCTAATGCAAGAT CTGGGGCAAGAATGTGTTCGGGTCAAATG AGTTA
BaFNuc	Fanzor2	MKRTYSATKSSLTLWTAASVKTT SAPKVVTTFSGWMKKILPTRAET SLTLINPADIADPSPPKKKAKKTTP ATPKPTLRIYKIGLRPSPAQRKTLN ACIVAANFAYNQCVHLVQHKVC KPHLYDLQKIVAKMKTPEDINHR YAPDRDGWFWKSSTIVRLLATKD FCAAYKAIVSNKKKDVAVIKYKTY DDPEAINPLSGLFGCQKQYATVT QAGLRLLPRLFGKDPIPLVKKKLK VATIDHDFKIEKTSKGKFVLCLTVE CSLLRRVKPPAPLFEDGYIHACGI DPGVRSFVTVYDPTRQDCYQFG TSAQKAERLDPITNAIDNWNSFV	CGCTCGAGTGGAGGGAACTGACAACTGTA GAGTGGAGATCACCGACGAACGCTGGACC TCAAAGACGTGTGGCATGTGCAGATCCAT CCACCGCGAACTTGGGGCAAAAGAACTAT TCGAGTGCCCCAACTGCCACTACACCTGCC ACAGAGACGTGCACGGAGCTCGAAACATC TTGCTGAGAAGCTTCGGACAGTTTCCAGTC TAGAAAAACACGAACTTTTTCCTTGGCCCA AGGATTGCCAAACACCACTCAAATCCTCGT TCAGGGGCCTCGAGTGGCTGGGCATATAT GGGTTA

		DQHRDKAPPTAIESWSRKTKKL WYKLKNQVRSLHDQVIAHLLGA YNFISLGKLDVSCFRRGTTAKSTN RWLRIYRHFEFRTKLLARVEGTD NCRVEITDERWTSKTCGMCRSIH RELGAKELFECPNCHYTCHRDVH GARNILLRSFGQFPV	
KnFnuc	Fanzor1	MDEGADDSEEAKRKRPDITLRRA LRKDKETSVVQTGWKFLCQELGI RDRIEEIIPEVTRIRVETCLLLNLHFI RLLDEGRPIPVIDQNLVGRAMQC TYSKNPQADPDLHETFVHHYLPL CPNRPNNSCLPRITNVLLDLRNQ LLSNIKNHVAVLFQSRHRAFMKL LLREAAPDVPFFGDADEDLESCT RLLTTATLWRPNESVRELLPEYPR IYGRIPEAAIECLQDLVDSVRPEV GPLPAAPQSRPHLYMPWMRIIS EEFSDRELRSFSLVPHASFSAPFIA ITPTTWPELQPKSGKRKAPGELR DAFPSIGRLESGGKTFADRITTDG VSASVYFLVEKRTPPPEDRVVHIH PKQRVVGLDPGKHPDFLTGIAVT GDWDGIERQEEIIGLGTRDFYHR AGFKKRTFLMHSWMSRDLDVA AFNKDAPSGNTVSLEDFGKRVTF VCANLYVLVRFHTARRVRKLRRR VTIKKQIEVDRACKRITAGKKTVV AFGAAQVWAGRTKRQCGPCES VKRRLSSHHKATVVMIDEFRTSQ VCSTCHSDVGKFAVLKRQRVME DGLPTVTEGGRREDEDEDGGGR TSYKTCHNVRACTNPLCRMVWN RDVNAARNIAWICMSIARGEGR PAEFTRAGVWG*	GGAGGAGGGAGGACAAGCTACAAGACGT GCCACAACGTGCGAGCGTGTACGAACCCG CTCTGTCGCATGGTGTGGAATAGAGACGT CAACGCAGCTCGTAACATAGCTTGGATCT GTATGAGCATAGTCAGAGGCGAGGGCAG GCCAGCGGAGTTCACGAGAGCAGGAATG TGAGGATGACTGAGAATTAGTCGAAAGAC ATAGCTGCCTAGAAACGAGTTCATCTAGG CACTTCGGTGAGAATCCGAGATACGGCTG GGTACTGTGGCGAGTGTGCCATTTTACTCT GAACAGACTGTA
CrFNuc	Fanzor1	MAPKRRRDEAEKAEEEKDHTTST KCGLAGLLSEKIEADGVAVTREES LAAVDFLVAALTRLRFEALCLLGL VAVRMCEDARREGQGLQPHCA TCRRLRKTELVEDDMYAAICAVS VCDLTEQGRKRGRPSKRDQHPE DDLFRHVCEEHFPRDEEAAGARV	GCCGCCATGGCCGCCGGCGGCGGCGGGG CCGGGCTGAGAGCCTGAACGGCGCTAGCA GGGCGTGGGGCTGAGGGTGCACGTGTTG ATTGGCGGCGAGTGACGTGACTAGTTTGT TAGCTGCGGGTTAGCACGGACTGTGCACC CCACCCCACCGGCCACGTTCCGGATTTGCG GGGATGCAAAGGCCCCCAACATAGAGGC

		NRSGLTPFLPPLSKGVFTNVKNH YAANFAAWLARSFRCRIDDELRE LRTPATKKLDKLAWSMAHAVLY DGELEQPRWWVGWAQGAAGA AAAAAAQGAGPAGGAAAAQA WTALVDYVNAQRASKRAAELLL REVKGAQATYKKASTRHMEWA AEILAGLEARRDQLGAQVQQLT QAQPLTREDTQRLASLRRELHRA RPFTLTPSPSFAPIYVPLDNTSMA RLPGLLPTLARRHGEVFAGAGAG AVAPSSFVQAAFGGGGMQSSAT LNAVGWGLFQLGGVTSRNAPFA NYITTDGVACSVAREAHNKPLAN LKPATAPADAEELCTLEEMKATQI IGVDPCGGGNWFMAARSPLYQ PGPWAWEGVGPAQRYLLELHD KQLDEELFPGQLPPEPRRRRKGV HRRKQSKHWQPRARTARRRRQ KRGRFHMSMGHWRHMSGLER LQPNRPQLAPALQAYVGGIPTAA TASAARFEERLRYLFASGAAGQA AGGPAEAGPRGAVHVLWHYHF SAFRRKRWAAFIQRDRALHRVA KQLTGGRPKEEVVVGWGSWAF QGGKGGSPISVRGGRAPTGRLIK LLRERYAKHVFIIDEYKTSKTCYNC GCQEMAIKRLGGLKEGQRPWSV KVCNDCLTTWNRDVSAANVIRV LLLLKLMGFERPTKLQRPPWPPA AAGPG*	GTGTGCTTAGTAGGCGCCCGCGTCAAGGT GGCTGGGTTGATAACGACCCGGGAGGGG AGGGCTCAGCCCTTTTCCTGCCTCCCTAAG GCAGCCACCTCCTTGT
TvoTnpB	TnpB2	MKRANAVKLIVGKETHEKLKELAI VAAKCWNEVNWLRMQQFKEG ERVDFSKTEKEVYEKYKQILKVNT QQVARKNAESWRSFFSLIEEKKG KLPKWFKPRPPGYWKDKSGKYK MLIIIRNDRYEIDEEKRIIYLKDFKL SLSFNGKLKWRGKQGRLEIIYNEA RRSWYAYIPVEVQNDVKAEDKLK ASIDLGIINLATVYVEDGSWYIFK GGSVLSQYEYYSKRISVAQKTLAR HKQGRSREMKLLHEKRKRFLKHA LNSMVRKIMEEFKNKGVGEIAIG	gggaagcccatgatgatgggcgtattaagcgtggtctc tataggtgtctccgcatagggaaggtaataaacgcaga cctgaatggtgcaataaatatcctacatatccccgagtc cctaggagctgggagcagagggcaactcacagtgagg gataggggtaatgggctgaagacccagcccgcggtct accgctggacgaatggagcggggtgggtgtcctcaccc actagctatgaagtgatgaaaatgaaggcggtaaact gcaaaccaatgaatcgccacaagggaaccttcacccctt tagg

		YPKEISKDHGNKLTVNFWNYGYII RRFEGVGEELGVKVVKVDEAWT SKTCSLCGEAHDDGRIKRGLYRCL RIGKVINADLNGAINILHIPESLGA GSRGQLTVRDRGNGLKTQPAVY RWTNGAGWVSSPTSYEVMKMK AVNCKPMNRHKGTFTL	
Isdra2 TnpB	TnpB1	MIRNKAFVVRLYPNAAQTELINR TLGSARFVYNHFLARRIAAYKESG KGLTYGQTSSELTLLKQAEETSWL SEVDKFALQNSLKNLETAYKNFFR TVKQSGKKVGFPRFRKKRTGESY RTQFTNNNIQIGEGRLKLPKLGW VKTKGQQDIQGKILNVTVRRIHE GHYEASVLCEVEIPYLPAAPKFAA GVDVGIKDFAIVTDGVRFKHEQN PKYYRSTLKRLRKAQQTLSRRKKG SARYGKAKTKLARIHKRIVNKRQ DFLHKLTTSLVREYEIIGTGHLKPD NMRKNRRLALSISDAGWGEFIR QLEYKAAWYGRLVSKVSEYFPSS QLCHDCGFKNPEVKNLAVRTWT CPNCGETHDRDENAALNIRREAL VAAGISDTLNAHGGYVRPASAG NGLRSENHATLVV	GATTCAAGAATCCCGAAGTGAAGAATCTT GCCGTCCGTACATGGACTTGCCCGAACTGT GGGGAAACCCATGACCGAGACGAGAACG CTGCGCTGAACATTCGGCGTGAAGCGTTG GTGGCTGCGGGAATCTCAGACACCTTAAA CGCTCATGGAGGCTATGTCAGACCTGCTTC GGCGGGCAATGGTCTGCGAAGTGAGAAT CACGCGACTTTAGTCGTGTGAGGTTCAA

Table S3: NLS sequences used in the study.

organism	family	NLS Sequence
Cotonvirus CTV1	family5	ATGGACTGTTTTATCACTTGCTTGCAGTCTTGGGAGAGAATTTTGAAACGAAAGCAACAGAAGAAAAGGCCG CGCTTGTTCTCTATTCTCCCTCGGAAGTCTGGATTCACTATAAGCTATGTCCCAATATCAAGTATGACATTGA TGAAACTTCTGTCCATGGGGGATAGTCCACTTGAGTCTGTTAGAGGGGACGGACGACATGAGAACCACAGC ATGCTGTGGCGGAAGTATTTCAATGTGCAGGGTTTGGAGACAAAGGGAAGTCGCTTCGATAATCGCATCTT GTCTGACGGGAAA
Prototheca cutis	unclassified	ATGATGAGGGAAGTTTCTAAAAAAGGGAAAGGAAAGGAAAAGTCCTCTGCTTCCACTTCAAGGAGTAGGAA GAGGAAGAGGAAAAGGCAAAAAAGGTCTTCACAAGCTGCCTCTTCTGCCAAAGCCAAGATTACCGAAGGGA ATATGCCACAAATGCTCAGTTTGGTTAAGCCGTTGTTCGAGAAACACTCAATCCCGGTTTGTGGCGTAGACC CAGGGATTCGCACAACTGGGGTCTCCACTAAACTTCGGCACGACATCTATAACACATGGACCCTGAGAGCG TCCGCAGTTAATCAC
Andricus curvator	unclassified	ATGATGGCCTGTAAAATTGGCGCTCTGAAAAGGCGCAAGGGTAAACACGGTAAGATTAATATAAGCTATGCG GAATACAAGGAAAATCCGTTCAGTTGTTGGAACTATGTTTTTGACATGTATAAGATTAAAAAACTGGGGAAAG GCAAGGAATTTGAGTTCCTGGCCCTCACCGATAATGTTGCAGTCAGCCTGGTCTATATTGAGCCCATTAAGC CAGAGAGCATAGGGCTGAGCTGGGAGCGGATTCGCCACATGTTTGAGGTACTCAAGTGCTTTGTTTATGAA ATAGGCATAGAT
Torulaspora delbrueckii	family5	ATGATGACGGAGATCAACTATTACTGGTTTAAAAAGAAAAAAAAAAAAAACATTGAGTCTAACTCTTGGTTTA ACATCAATAGCATAGAAAACAAGAAAAAAGAGTTTGAAGAGAATGATATACCTCTTAACAAATATCTGAGTTG TAACCAGATAGAAATTTACCCCAACAACTACCAGAAGGATATTCTTCTGAAATGGATGGATCTTTTCATAGAT ATGTATAATCACACAAATTATTTCATCAATAACAACATATACGACTTTACCAATAGGAAGATTAAGAACAATGT GAAAGAC
Globisporangium ultimum	unclassified	ATGAAACGCAAACAGCAGAAGAAACGACCGAGACTCTTTTCCATCCTTCCGCGCAAGTCAGGATTCACCATT TCCTACGTCCCTATTTCTAGTATGACACTGATGAAACTGCTTTCTATGGGGGATAGTCCTCTTGAAAGCTTTC GAGGGGATGGCAGACACGAAAACCATAGCATGCTTTGGAGAAAGTATTTCAATGTGCAAGGATTGGAAACC AAGGGCAGCCGGTTCGATAATAGGATCTTGTCCGATGGTAAAGGCGTATCAGTCCAGATGAAGCATCAGGC ATCAGAGGACGTG
Globisporangium ultimum	family4	ATGATGATTAAAGAAAAGTACTCTAGCAACAAGCGCAAAAGGTATCCTACCACACACCGAAAGAAACGCATG TCAGACGCCCAAATCAGTACGAAAGCTACGACAATACACGGCAGAAGCATCCCTCATACTCAAAAGAGCAA GCCGCGAGCAACGCCGCCCATATCAGTGACATATGAGACCTTTTGGGGCTTCGACCCTGGTAGAAAGGCTA TGGTCGCCGGGGTGGTAAGAACCGGCGACCCGATGCCTTTCAAAGACGACAAGACACGGGATAAGCCGTT TTATGTGCGGAGGTCA
Scenedesmus sp. PABB004	family1	ATGATGAATGAAATCCAACTCCCTACCCCGAGGGGGTCCGCGAGGCGGAAACGAAAGAGACAAACCGAAC CCCAAATAAGTTACGATCAGGCCAAAAACACTTTGCTTGGTGTGCTTTTGCAGAAACTCAAGTCCACAATCG CTAATGAACTGTGGAAAAACGAGACGAAGAATAGTCGGAAAACAAAAAAAGGTAAGCAGTCAGCCAAGACT AAGGGCAAGCGCTTGAAAGCACAAGAGAAACGGATGAATAGACTTAGACAGGCATTGGAAAAGGCTGACCG CATCCCCCGGGGCAGT
Scenedesmus sp. PABB004	family1	ATGATGAGCTATGGGATTGAGATTGAGACGGTAGCAAAACGAACGAGCAAAAGTAAAAAAAAACGGAAGTT CGCACAGCAACTGCATTCAGATGGAGAAAGCGTTACCATCCTGTATGAGTCAGAGCTGAAGCCCGTCCTGA ACGTAGAAAATAAAATCAGAAAAATTAAGAAAAATTACGCACAGGATAAATACAAGTATCGAACAAGTATCGA TGTGGGATTTAATACATATATTGCTGCTAAGAGACTTACCATAGAGACTGGCAAGGAAGTCAACTTGAAACT CAAATCTAAACAT
Chlamydomonas sp. ICE-L	unclassified	ATGATGAAAGAGGCAGTGAAGAATGTGAAACCCAAAGTGCCAGCGAAGAAACGAATAATTACAGGTAGTAA AACTAAGAAGAAGGTTTTCGTGAAAAAGAAGCCGCCGGACAAAAAACCCTTGAAGAAACCCGTCAAAAAAAC AGTTAAAACAGATAAGCCCAAGTCTATATATGTCCCAAATAAGGATTTGAAAATTTCCAAATGGATACCGACA CCTAAAAAAGAGTTCACGGAAATAGAAACGAACTCATGGTACGAGCACCGCAAGTTTGAGAACCCCAACAA GAGCCCGGTCCAA
Chlamydomonas sp. ICE-L	unclassified	ATGCCTTTCCTCTGCACGACTCGATACTGTAGACGGCCAAGCAAGAATGAGAAAAGAAAGCGCAAGACCTC TCACATTTTGGTGGTGGCACTCCAAATTTGGATTCATAGCGTCGCTCATAGTGATTTTGAGAATGGACTCTTT TTGTATGGACAAAGAGTGCCGCCTAGCATACACGCAAGGCAGACTTTTGCTTACAAGAATCACCAAGAATAT GGGTCCGCCAGCCAAACGAGTCGAAAGTGGCGCCAATTCCTTGGGTCCTCAGTGACGAGAATTGTAACTGA TGGTTTCGCCGTC
Chlamydomonas sp. ICE-L	family4	ATGAAGCGAGCAGGCGGTCGAAAAGGAGGTCACCGGCGAAAGCAGTCAAAGCATTGGCAACCGCGGGCA CGAACCGCAAGAAGAAGACGCCAAAAAAGAGGAAGACTGCACATGTCCATGGGCCACTGGAGGCATATGA GTGGGCTTGAAAGACTGCAACCAAACAGACCACAGCTCGCACCAGCGTTGCAAGCGTATGTCGGAGGTATC CCAACTGCCGCTACAGCATCATCTGCCCAGTTTGAAGAGAGATTGCGCTACTTGTTCGCCTCTGGCGCTGC GGGGCAGGCTGCCGGACAG
Chlamydomonas sp. ICE-L	unclassified	ATGATGCGGGAGGTCAAGGCGGGAACTAAGAGAGCGAGACAGCCTGAGGTGAAGAGTGTAGCATTGAAAA AAGCTAAGAAGACAGGTAGGGCTTCCAAGCAGGCTTCTTCCTCTAACACGGCGTTTATAGCCAATCCCTGC CCTGCCTGCGGGAGCACTGAACACAGATCAACGCGAGCGAAGTCTTGTCCCCGGCATACTAAATCAACGGA AGAAATCCTTGAGGACCAACTTGGGAAGGGCTTTGAGACATTTACACGCAAAGCGCCTTTGCACTCTGTCGT AGTCGAAGCACACAGA

Catovirus CTV1	family5	ATGTACCTCTTGATGAAGAAGAAAAAAGAACCTGACAAAAACAAAAGTGACAAAGAAAAAGAGTATGAAGAA AAGTATCGAAAGTATATCACATCCTATAAGACACACAAGACATCACTCGAAAACATTATTAAGAATACCGACG ACTTGGAAAAAATCAACGACGCGGTTTATAGGATTAACCTTATCATTTGCCACACCTATCAGTTCCTTAAACT GTACTACCTGTACGAATTTCATAATAAGGGTAGAATTATAGTGATTGATGAGCAACTCGTGAACACCACGAT GAAGTTGTTC
Indivirus ILV1	family4	ATGATGAAAAAGCCTAAGGTGAAAGAGAAAGAGAAGGAAAAGGAGAAAGAAAATTTCGATTTTATGAAGACT AATAAGGGGAATATCCATAAGCTCATAAAGGATAAGATGGTACTCTCTATAATCGACGAGTTGGTGGTGAGG GTCAACAAAATCGTAATCCATGCGTATCAGTTTAACAAACTCTTTTGCTCCTACCTGTATAACAACCACTTTC CTCTTCCGTTTCTGGACAAAGAGTACATTTGCGACATATTCCGGGTTATCACGAAAAGAAAATGCGGTAAAG GAGGTTATACT
Apophysomyces variabilis	family1	ATGATGGAGACTATCGTAAATAAAGAACCACCCGACAAGCGCACCCGCCCGGATCGGGCTGCAAAAATTAA AGACCGCAAAAATGGGGAAGAAAACGTCGTTAAATGTACTCTTTCCAGGATCATAGCCGATAAGGACAACAA AGACGTAATTACCAAAGTCATCAAAGAGAGGGTGGACAGTGTATCCAAACGAACGTATTTGGCCGGAGTAG CGGTGAACAGGATTTTGAAAGAAGTCTTCGAGGGGGTACCTGATGCTCAGCTCCATACGGTCTCAGTACCT TCCGTTGACCAAAAT
Apophysomyces variabilis	unclassified	ATGAGCCCCGGATCATCTGCGGCGAGAAAGAAAAACGAGAAGCAGTGTCGGGTGCAGAAGAAGCGAAAGA GACGCGGCCCGAAAGGTGGGGGTCCGGCCAGTAAAACCGCAAGAAAGACGACAGTAAGGAGAGATGGGG GGGAATCACTTTCTTCTTGGACGTACCAGGGTGTAGTTCAAGGGAAAGTAGCTAGGAGATATAGAAAAACAT GCCGGTTTGTTCATCAGGAACACCATAAGGATGTGGTGCGAACAGACGCTTTCGGGGAGAGGGATGTCTCA AGAAGGGATGCCCATGGG
Apophysomyces variabilis	family4	ATGATGGCAAGCCGAAACAAGCGGAAAAAAAAGCCGCAGGCGAGCACGAGTGCCGACACCCAGAGCGAC GACGATTTCCAACAACTCCTTCCGCCGAAGGGTAAATTGAATATGAAAATGCAGATGACGGATAGTGAGCC GGGGCCTTCAACCCTCCCGTCCTCCAGTATGGTCTCTACGCAAAATAACAGTAGCAAGGAAAAGCCCCGGA GTCTTAAAGGGAAAGAAAAGGCGACATTCGACAGTGACGATGATTTCCAACCCTTGCCCCCACCAAAACGA AAACAACCTCTGCAGGTT
Apophysomyces variabilis	unclassified	ATGGTTCACCTTATACTCATTCTTATGACGAAGAAAAAGAAAAAATTCAAGAAAAAGAAGATTTTTTACAAAAA ATACCACAAATTCAACTGGCTCTCCAGGCTCTTCAATGATAATCAGTTTAGTGATCTTAGTTCCTATATATTCT TTCATGGAGACCATTTCTGGAAGCACAAGTATAACTTGCCATTCAAGAGTCAATTGTCCTTTTCATGGTTTGA TTTTCATAAATCCATACAAATTTCTCCAGCAAATCCTATAATGCCGTCCGATGCCAGGTTCGTGGACAAGACA AAATTT
Cyanidioschyzon merolae	unclassified	ATGCCACTGACGCGAAGGCGACGACAAAAATCCCGGAGAGGGCTTCACCGGAGACATAGGACGAGGCGG GCGCGACGCAAAGAGCGAGTCATCGAAATCTCCACCCCCAAGTATCGACATCTCGCCCGGATGAACGACTT CAGGTTTTACAACGAAAATCTGAAGATGAGAGAGCCGTGGTATGCTGGCGTTATGCGCGCCATGCCAAGTT TCAAGACAAGTAGCTATGATCTTTATTTTCAGAGGCTTCAGTTTTTCTGGAAGCACCTTAGATTTCTTTTGGT GTTTAGTGCGGAACAG
Cyanidioschyzon merolae	unclassified	ATGTCTCCACGGCCGCAGCCGGCTGCGCCTCCTGCAGCGCAGGGAAGAGCCCGCGGGGGCGCCCCGGC ACCCGCTGGCAGACGAGGGGGGGCTGCAGCACCTAGACCGGGGGCGAGGAGACGGGCAGGGAGGGGC GCCGGTCCAGGTCCTGGTGCGGCTCTTCCCGCGGCGGCTGCCGAAAGACTGCGCCTGGCCCTTGCCGCA CTTACAGCTGCCGGAGACGCAACAGTCGGAGCCGACCTCGCTGCGGTCGAAGCTGAGTGGCAGGCTGCC ATGGCGCAACATCACGAACGGGACGCC
Chlamydomonas reinhardtii	unclassified	ATGTGTAGGAGGTGCCGCATCACGCCACTTTGGCTGGCTGGTCGGAGGATGAAAAAGAGACGACGACGGGTCCTCCG ACCCAAAAAGTGCATGATAACAACCCTGTCTCTGGCCAGAACACGGGGTATCCGAATAAGTTCTCCCCTGCAGCATAGG AGTGTTCCCAAGATGGGCCTGACCGCTACTTATGTCAAAAAAGATATGCGCCTGCTGCGAATTTCTCGCGCCCGCTATTA CCGGGAGAGCGGAATAATGAACGCGTGGAAAAAAATCAGCGCCTGGGATGCGGGCATGAAGGAC
Contarinia nasturtii	family3	ATGATGTATTGTATGCATGAGGATTCTAGTCATAAAAAGGGTCGGCGGCGGACGATGCGGATCAGCTCAAGGGAGTG GGCTTTTCTGACTCGATCTCGCAAATTTCGACGCCTGTTGAGAAGGCTTATGACCAAAGGCGTAAGAAAGGCATTGCAA CGACTCGCGGCGGAAGCACCGAATGGGGCTCGGGCGTGCACCGTGGCGGACTTTCTTGCTTACCTCAGAGTCTTTGGG CAGGTAGAGGCCACTTTGAGCAAGCATTACAGTAAACCCATCTTCAGAAAACTTAGGCTGTGGACG

Table S4: NGS primers used in this study.

NGS Primers	Name
ACACTCTTTCCCTACACGACGCTCTTCCGATCTCtggaattgtgagcggataacaatttcacacagg	TAM_NGS_F1
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTctgcaaggcgattaagttgggtaacgcc	TAM_NGS_R
ACACTCTTTCCCTACACGACGCTCTTCCGATCTCacgtggagtccaaccctggacc	Luciferase_Indel_NGS_F1
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTtcagcatcgagatccgtggtcgc	Luciferase_Indel_NGS_R1
ACACTCTTTCCCTACACGACGCTCTTCCGATCTctttgttggagttcgttttcttccttgaaattgttgg	EMX1_Fanzor2_NGS_F
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTattgactgtagacctagactacagaccgtcac	EMX1_Fanzor2_NGS_R
ACACTCTTTCCCTACACGACGCTCTTCCGATCTCgggtcacagggcaagactttgtctc	HPRT1_Fanzor2_NGS_F
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTtgccaccacgcctggctaatt	HPRT1_Fanzor2_NGS_R
ACACTCTTTCCCTACACGACGCTCTTCCGATCTCatcattccaccaatcaggactcggc	dync1h1_NGS_F
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTccagcctggtcaacctagcgaga	dync1h1_NGS_R
ACACTCTTTCCCTACACGACGCTCTTCCGATCTCccttctccccacagcctccc	b2m_NGS_F
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTgctgtaaactagccaggttgggaatatattgcc	b2m_NGS_R
ACACTCTTTCCCTACACGACGCTCTTCCGATCTCgtctgagtcttcaagttttcactccagctaacac	cxcr4_NGS_F
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTacagtcctaccacgagacatacagcaac	cxcr4_NGS_R
ACACTCTTTCCCTACACGACGCTCTTCCGATCTCagagactcagagtccaagagggaagcc	CA2_NGS_F
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTactagggagtggcttatgcacaggtatattatgtg	CA2_NGS_R
ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTCCTTCAGTTCTATCCATGTTGTTGCAAATGGTAAG	DMD_NGS_F
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTCTTTAATCAATGCTTTGTAGTTTTCACTGTATAAATATTTCACC	DMD_NGS_R
ACACTCTTTCCCTACACGACGCTCTTCCGATCTCATGTCTGGAATTGAGCCAGGTACTGGG	Grin2b_NGS_F
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCATTAACCAGGTACTGGCCCACTATAGGG	Grin2b_NGS_R

Table S5. TAM library and spacer sequences used in study.

Name	Sequence
TAM Library Plasmid	gatcaaaggatcttcttgagatccttttttttctgcgcgtaatctgctgcttgcaaacaaaaaaa ccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaa ctggcttcagcagagcgcagataccaaatactgttcttctagtgtagccgtagttaggccacca cttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgct gccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgc agcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcgaacgacctacaccga actgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggac aggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacg cctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatg ctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcctggcc ttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgta ttaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagt gagcgaggaagcggaagagcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcat taatgcagctggcacgacaggtttcccgactggaaagcgggcagtgagcgcaacgcaattaatg tgagttagctcactcattaggcaccccaggctttacactttatgcttccggctcgtatgttgtg TGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTT NNNNNNNGCAGCCACCTCCTTGTTATTGGGTACCGAGCTCGAATTCACTGGCCGTCGTTTTACA ACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGcacatccccctttc gccagctggcgtaatagcgaagaggcccgcaccgatcgcccttcccaacagttgcgcagcctga atggcgaatggcgcctgatgcggtattttctccttacgcaTCTGTGCGGTATTTCACACCGCAT ATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCGCCA ACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGA CCGTCTCcgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgagacgaaa gggcctcgtgatacgcctatttttataggttaatgtcatgataataatggtttcttagacgtca ggtggcacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaa atatgtatccgctcatgagacaataaccctgataaatgcttcaataatattgaaaaaggaagag tatgagtattcaacatttccgtgtcgcccttattccctttttgcggcattttgccttcctgtt tttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgg gttacatcgaactggatctcaacagcggtaagatccttgagagttttcgccccgaagaacgttt tccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgccggg caagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtca cagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataaccatgag tgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgctttt ttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagcca taccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgttgcgcaaactatt aactggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcggataaa gttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggag ccggtgagcgtgggtctcgcggtatcattgcagcactggggccagatggtaagccctcccgtat cgtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgag ataggtgcctcactgattaagcattggtaactgtcagaccaagtttactcatatatactttaga ttgatttaaaacttcatttttaatttaaaaggatctaggtgaagatccttttgataatctcat gaccaaaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaa
21 nt report	GCAGCCACCTCCTTGTTATTG

er Guide sequence	
EMX1-1	aaaaaaaagaaaagaaaaaa
EMX1-2	aagagtggccttgatttgta
EMX1-3	aaataaaatttaaaaaaaaa
EMX1-4	gtttccagttttattttgtta
EMX1-5	gagaaacaaatgaaagggac
DYNC1h1_G1	gagatggtaggttcttctaa
DYNC1h1_G2	aatacacatagatatagggtc
DYNC1h1_G3	aaaaaaacaaaaaaaccaaaa
DYNC1h1_G4	aacatcaaagtgcactgtcag
DYNC	caaaattcttaattt

1h1_G5	
B2m_G1	gtgatcatgtaccctgaata
B2m_G2	aaagaattttatacacata
B2m_G3	tacacatatatttagtgtca
B2m_G4	gtagcactaacacttctctt
B2m_G5	aatacacttatattcagggt
cxcr4_G1	tatctgaaaaatgtgtaact
cxcr4_G2	tacgataaataacttttt
cxcr4_G3	agttacacatttttcagata
cxcr4_G4	tatctgaaaaatgtgtaact
cxcr4_G5	attgacttatttatataaat
CA2_G1	tagtcagaagaagaagtttg
CA2_G2	cagaaagatccaaacttctt
CA2_G3	ttcatctgacaacttccttt

CA2_G4	tagatgaggagacttgtaga
CA2_G5	attctacaatgatatattgt
DMD_G1	TATAAATGAATATTCCGTTGT
DMD_G2	TCCATTTATCTGTTAATGGC
DMD_G3	CAGTATCATCAGGAAGAATAA
DMD_G4	TTCTTCCTGATGATACTGTA
DMD_G5	GTTAAATTTATTCCTCTTTT
GRIN2b_G1	GCTCCCTAAGGGGACAGACC
GRIN2b_G2	AGTTTAACTTTATGAAATTGC
GRIN2b_G3	ACTTTATGAAATTGCCTTTT
GRIN2b_G4	TTATATGTCAATAATGGTTA
GRIN2b_G5	TATGTCAATAATGGTTATTTC

Data S1.

Table of all discovered Fanzor sequences in this study.