Science Advances

AAAS

advances.sciencemag.org/cgi/content/full/3/8/eaao4774/DC1

Supplementary Materials for

Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity

Alexis C. Komor, Kevin T. Zhao, Michael S. Packer, Nicole M. Gaudelli, Amanda L. Waterbury, Luke W. Koblan, Y. Bill Kim, Ahmed H. Badran, David R. Liu

Published 30 August 2017, *Sci. Adv.* **3**, eaao4774 (2017)
DOI: 10.1126/sciadv.aao4774

This PDF file includes:

- table S6. Base editing outcomes from treatment with BE3, CDA1-BE3, AID-BE3, or APOBEC3G-BE3 at the *RNF2* locus.
- note S1. Python script to detect linkage disequilibrium in base editing outcomes at target sites with multiple target cytidines.
- Supplementary Sequences. Amino acid sequences of CDA1-BE3, AID-BE3, BE3-Gam, SaBE3-Gam BE4, BE4-Gam, SaBE4, and SaBE4-Gam.

SUPPLEMENTARY FIGURES

a

EMX1: GAGTC_5C_6GAGCAGAAGAAGAAGGG

FANCF: GGAAT$C_6C_7C_8$TTCTGCAGCACCTGG

HEK2: GAAC_4AC_6AAAGCATAGACTGCGGG

HEK3: GGCC_4C_5AGACTGAGCACGTGATGG

HEK4: GGCAC_5TGCGGCTGGAGGTCCGGG

RNF2: GTC_3ATC_6TTAGTCATTACCTGAGG

b

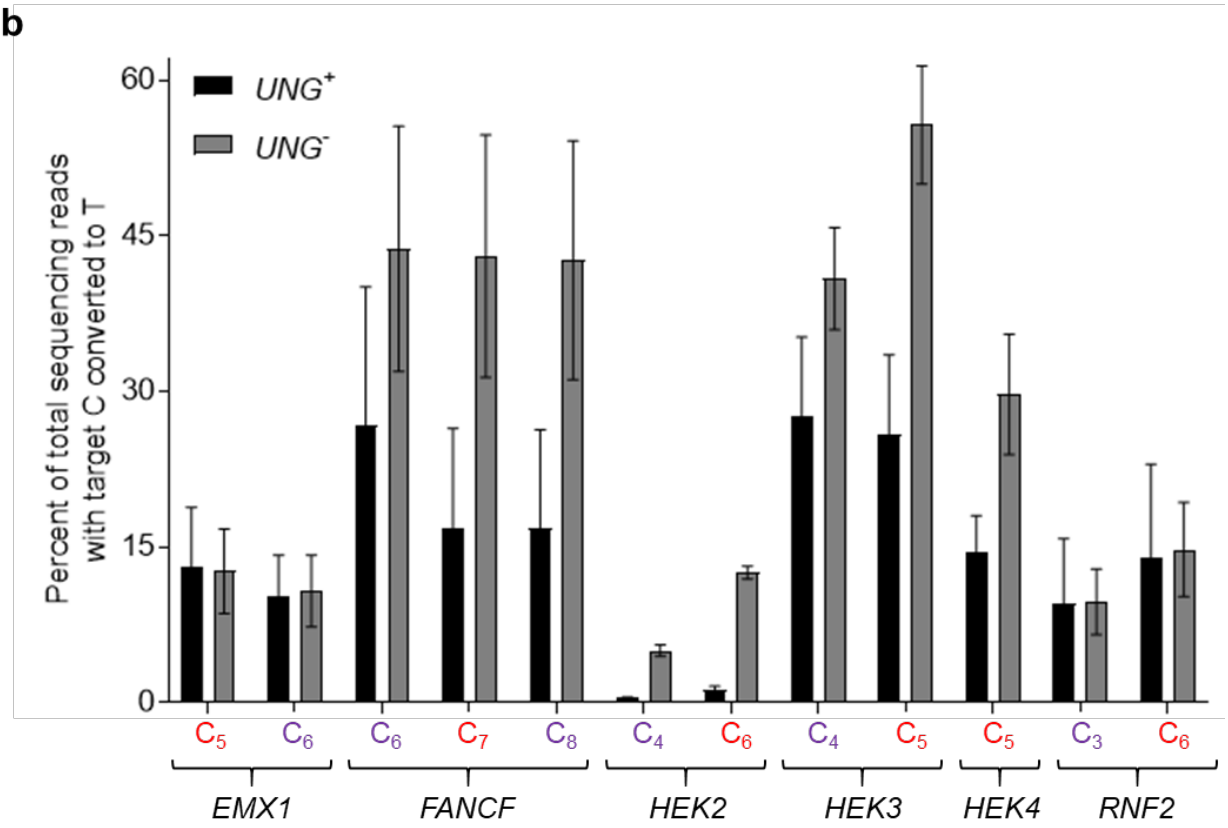

Figure S1. Base editing efficiencies in *UNG* knockout cells. (**a**) Protospacer and PAM (blue) sequences of genomic loci studied, with target Cs shown in red and purple. (**b**) HAP1 (*UNG*$^+$) and HAP1 *UNG*$^-$ cells were treated with BE3 as described in the Methods. C-to-T base editing efficiencies are shown.

a

EMX1: GAGTC_5C_6GAGCAGAAGAAGAAGGG

FANCF: GGAAT$C_6C_7C_8$TTCTGCAGCACCTGG

HEK2: GAAC_4AC_6AAAGCATAGACTGCGGG

HEK3: GG$C_3C_4C_5$AGACTGAGCACGTGATGG

HEK4: GGC_3AC_5TGC_8GGCTGGAGGTCCGGG

RNF2: GTC_3ATC_6TTAGTCATTACCTGAGG

c

Percent of cleanly edited reads with given sequence

HEK2 protospacer sequence	BE3	CDA1-BE3	AID-BE3
GAAT_4AT_6AAAGCATAGACTGCGGG	77 ± 1	88 ± 2	92 ± 2
GAAC_4AT_6AAAGCATAGACTGCGGG	23 ± 1	5 ± 1	4 ± 1
GAAT_4AC_6AAAGCATAGACTGCGGG	0 ± 0	7 ± 2	4 ± 2

b

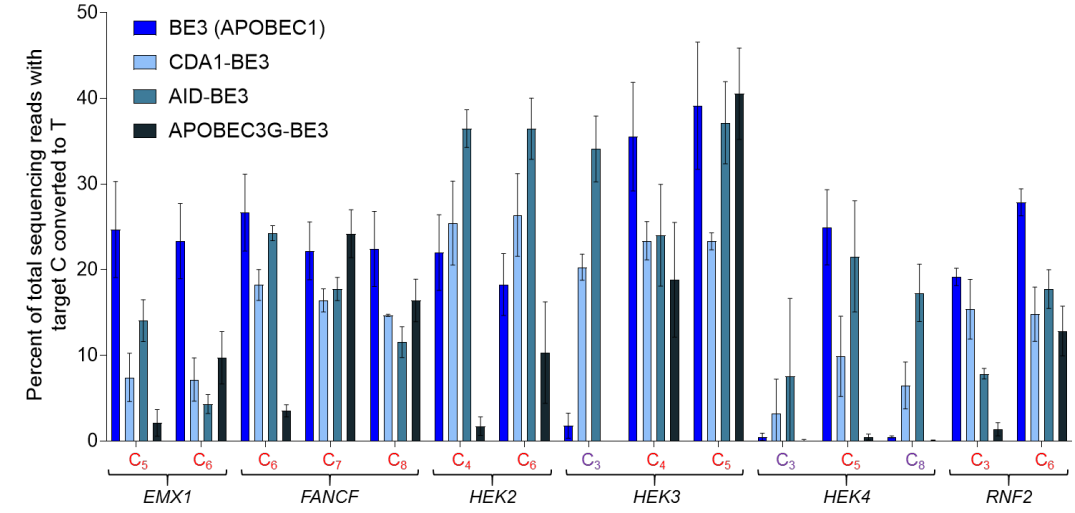

Figure S2. CDA1-BE3 and AID-BE3 edit Cs following target Gs more efficiently than BE3. (**a**) Protospacer and PAM (blue) sequences of genomic loci studied, with target Cs edited by BE3, CDA1-BE3, and AID-BE3 shown in red, and target Cs (following Gs) edited by CDA1-BE3 and AID-BE3 only shown in purple. (**b**) HEK293T cells were treated with BE3, CDA1-BE3, AID-BE3, or APOBEC3G-BE3 as described in the Methods. C-to-T base editing efficiencies are shown. (**c**) Individual DNA sequencing reads from HEK293T cells that were treated with BE3, CDA1-BE3, or AID-BE3 targeting the *HEK2* locus were binned according to the sequence of the protospacer and analyzed, revealing that > 85% of sequencing reads that have clean C to T edits by CDA1-BE3 and AID-BE3 have both Cs edited to T (Figure S1c).

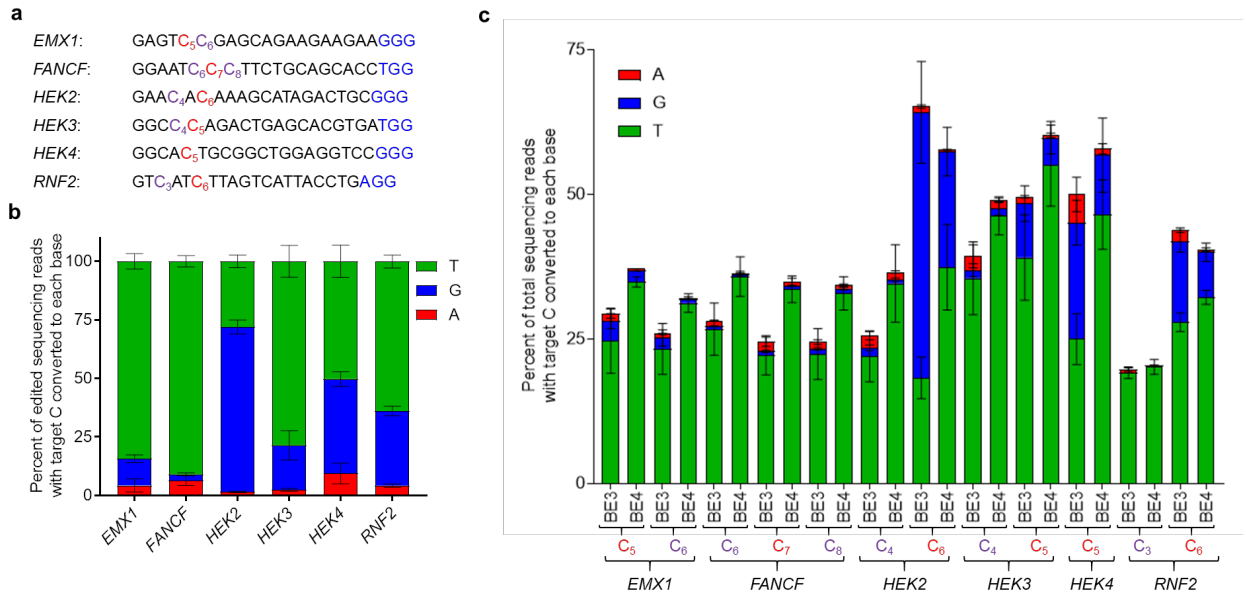

Figure S3. Uneven editing in sites with multiple editable Cs results in lower product purity. (**a**) Protospacers and PAM (blue) sequences of genomic loci studied, with the target Cs in (c) shown in purple and red, and target Cs in (b) shown in red. (**b, c**) HEK293T cells were treated with BE3 or BE4 as described in the Methods. The product distribution among edited DNA sequencing reads (reads in which the target C is mutated) is shown. C to non-T editing is more frequent when editing efficiencies are unequal for two Cs within the same locus. Values and error bars reflect the mean and s.d. of three independent biological replicates performed on different days.

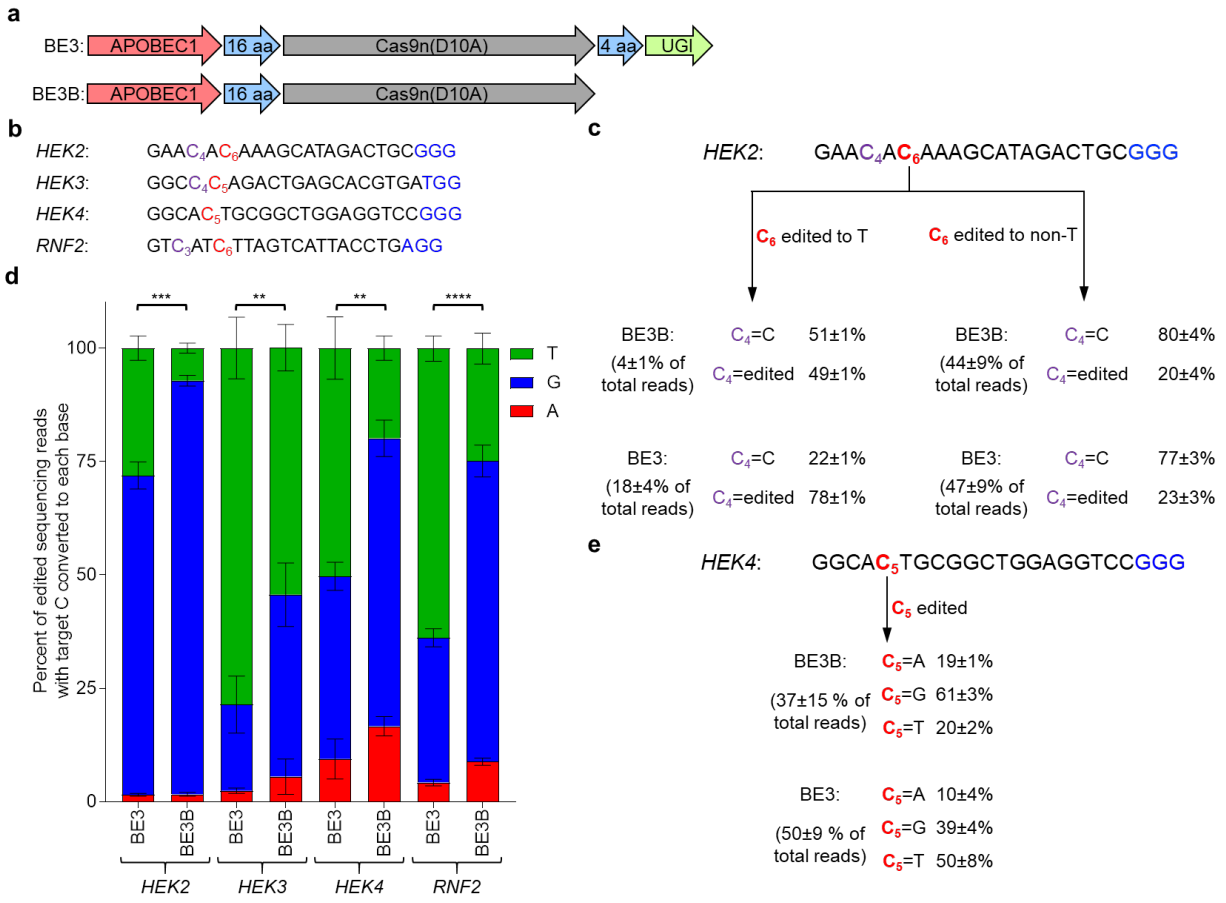

Figure S4. Base editing of multiple Cs results in higher base editing product purity. (**a**) Architectures of BE3 and BE3B. (**b**) Protospacers and PAM (blue) sequences of genomic loci studied, with the target Cs that are investigated in (b) shown in red. (**c**) The HTS reads from HEK293T cells that were treated with BE3 or BE3B (which lacks UGI) targeting the *HEK2* locus were binned according to the identity of the primary target C at position 6. The resulting reads were then analyzed for the identity of the base at the secondary target C at position 4. C_6 is more likely to be incorrectly edited to a non-T when there is only a single editing event in that read. (**d**) HEK293T cells were treated with BE3 or BE3B (which lacks UGI) as described in the Methods. The product distribution among edited DNA sequencing reads (reads in which the target C is mutated) is shown. (**e**) The distribution of edited reads with A, G, and T at C_5 in cells treated with BE3 or BE3B targeting the *HEK4* locus (a site with only a single editable C) show that single G:U mismatches are processed via UNG-initiated base excision repair to give a mixture of products. Values and error bars reflect the mean and s.d. of three independent biological replicates performed on different days; $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$ by two-tailed Student's *t*-test.

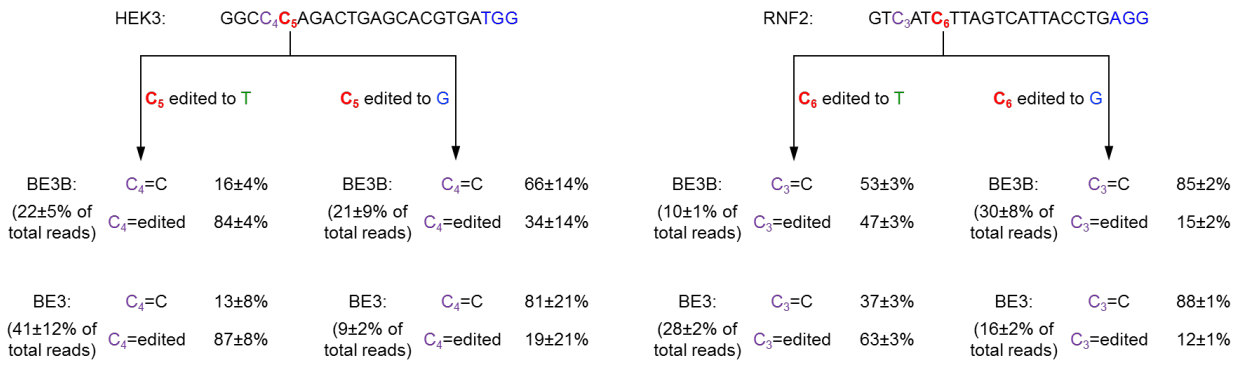

Figure S5. Base editing of multiple Cs results in higher base editing product purity at the *HEK3* and *RNF2* loci. DNA sequencing reads from HEK293T cells treated with BE3 or BE3B (without UGI) targeting the *HEK3* and*RNF2* loci were separated according to the identity of the base at the primary target C position (in red) The four groups of sequencing reads were then interrogated for the identity of the base at the secondary target C position (in purple). For BE3 when the primary target C (in red) is incorrectly edited to G, the secondary target C is more likely to remain C. Conversely, when the primary target C (in red) is converted to T, the secondary target C is more likely to also be edited to a T in the same sequencing read. These observations suggest that base editing product purity is impaired when only a single uracil intermediate is generated, perhaps reflecting more facile processing by UNG. Values and error bars reflect the mean and s.d. of three independent biological replicates performed on different days.

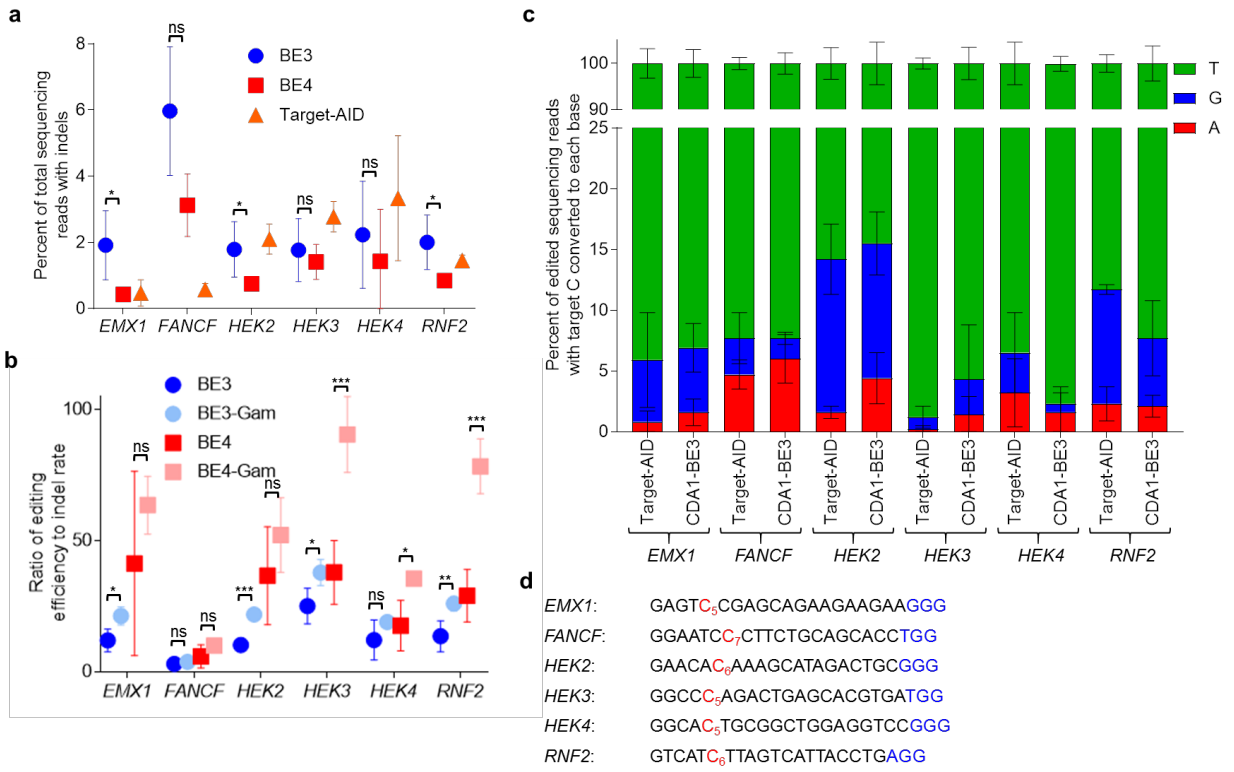

Figure S6. BE4 induces lower indel frequencies than BE3, and Target-AID exhibits similar product purities as CDA1-BE3. (**a**) HEK293T cells were treated with BE3, BE4, or Target-AID as described in the Methods. Frequency of indel formation (see Methods) is shown. (**b**) HEK293T cells were treated with BE3, BE3-Gam, BE4, or BE4-Gam as described in the Methods. The ratio of editing efficiency to indel rate is calculated by dividing the percent of total sequencing reads in which the target C (shown in red in Fig. 5b) is converted to T by the frequency of indel formation (see Methods). (**c**) HEK293T cells were treated with CDA1-BE3 or Target-AID as described in the Methods. The product distribution among edited DNA sequencing reads (reads in which the target C is mutated) is shown. (**d**) Protospacers and PAM (blue) sequences of genomic loci studied, with the target Cs that are investigated in (c) shown in red. Values and error bars reflect the mean and s.d. of three independent biological replicates performed on different days, except values and error bars of BE4 reflect the mean and s.d. of nine independent biological replicates performed on different days by two different researchers; ns: $P \geq 0.05$, *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$ by two-tailed Student's *t*-test.

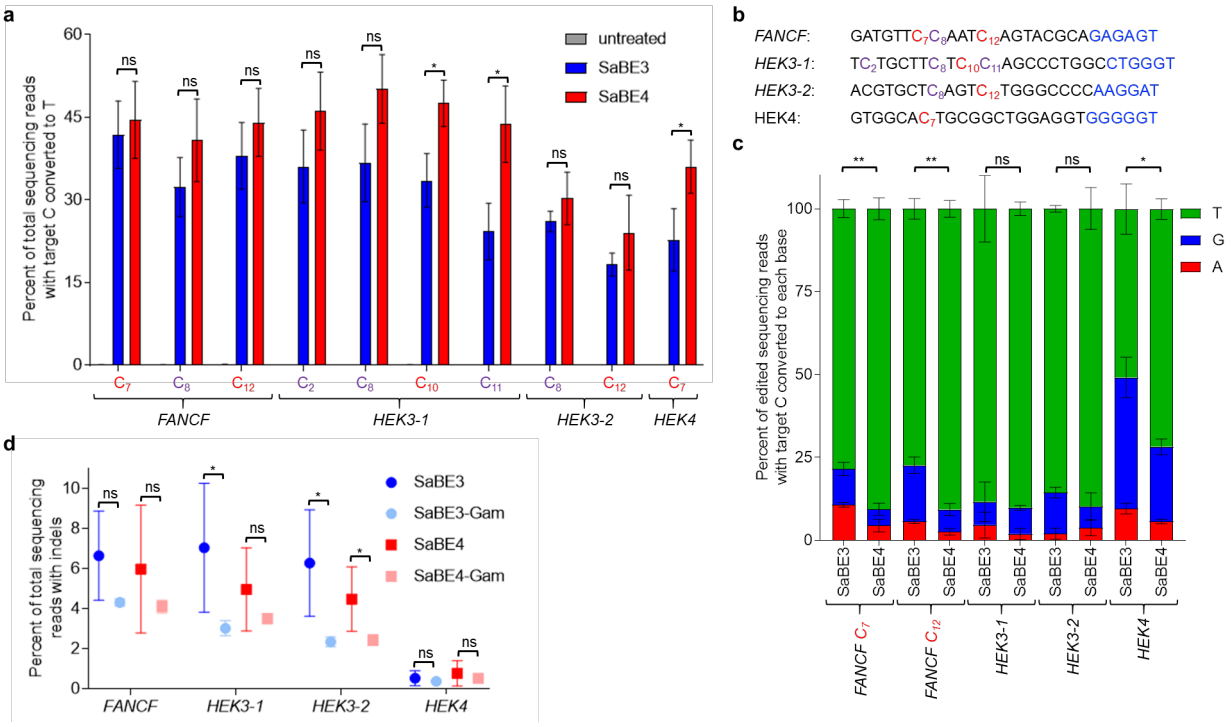

Figure S7. SaBE4 exhibits increased base editing yields and product purities compared to SaBE3. (**a**) HEK293T cells were treated with SaBE3 or SaBE4 as described in the Methods. The percentage of total DNA sequencing reads with Ts at the target positions indicated are shown. (**b**) Protospacers and PAM (blue) sequences of genomic loci studied, with the target Cs in (a) shown in purple and red, with target Cs that are investigated in (c) shown in red. (**c**) The product distribution among edited DNA sequencing reads (reads in which the target C is mutated) is shown. (**d**) Frequency of indel formation (see Methods) is shown. Values and error bars of SaBE3-Gam and SaBE4-Gam reflect the mean and s.d. of three independent biological replicates performed on different days. Values and error bars of SaBE3 and SaBE4 reflect the mean and s.d. of six independent biological replicates performed on different days by two different researchers; ns: $P \geq 0.05$, *$P < 0.05$, **$P < 0.01$ by two-tailed Student's *t*-test.

SUPPLEMENTARY TABLES

Table S1. Base editing outcomes from treatment with BE3, CDA1-BE3, AID-BE3, or APOBEC3G-BE3 at the *EMX1* locus. The sequence of the protospacer is shown at the top, with the PAM in blue and the target bases in red with a subscripted number indicating their positions within the protospacer. Underneath the sequence are the percentages of total sequencing reads with the corresponding base. Cells were treated as described in the Methods. Values shown are from one representative experiment.

EMX1		G	A	G	T	C_5	C_6	G	A	G	C_{10}	A	G	A	A	G	A	A	G	A	A	G	G	G
untreated	A	0.0	99.9	0.0	0.0	0.0	0.0	0.0	99.9	0.0	0.0	99.9	0.0	99.9	100.0	0.0	100.0	100.0	0.0	99.9	99.9	0.0	0.0	0.0
	C	0.0	0.0	0.0	0.0	99.9	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.0	0.0
	G	100.0	0.1	100.0	0.0	0.0	0.0	100.0	0.0	99.9	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	99.9	100.0	100.0
	T	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0
BE3	A	0.0	100.0	0.0	0.0	2.0	0.8	0.0	99.5	0.0	0.0	100.0	0.0	100.0	100.0	0.0	100.0	100.0	0.0	100.0	100.0	0.0	0.0	0.0
	C	0.0	0.0	0.0	0.0	72.6	76.2	0.0	0.5	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	G	100.0	0.0	100.0	0.0	3.0	1.1	100.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	100.0	100.0
	T	0.0	0.0	0.0	100.0	22.3	21.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
CDA1-BE3	A	0.1	100.0	0.0	0.0	0.2	0.0	0.1	99.9	0.0	0.0	100.0	0.0	100.0	100.0	0.0	100.0	100.0	0.0	100.0	100.0	0.0	0.0	0.0
	C	0.0	0.0	0.0	0.0	91.5	92.8	0.0	0.0	0.0	98.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	G	99.9	0.0	99.9	0.0	0.4	0.1	99.8	0.0	99.9	0.0	0.0	99.9	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	100.0	100.0
	T	0.0	0.0	0.0	100.0	7.9	7.1	0.0	0.0	0.0	1.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
AID-BE3	A	0.0	100.0	0.2	0.0	0.2	0.1	0.3	100.0	0.0	0.1	100.0	0.0	100.0	100.0	0.0	100.0	100.0	0.0	100.0	100.0	0.0	0.0	0.0
	C	0.0	0.0	0.0	0.0	85.8	95.2	0.0	0.0	0.3	94.5	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	G	99.9	0.0	99.8	0.0	0.7	0.0	99.7	0.0	99.6	0.3	0.0	99.7	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	100.0	99.9
	T	0.0	0.0	0.0	100.0	13.4	4.6	0.0	0.0	0.0	5.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
APOBEC3G-BE3	A	0.0	100.0	0.0	0.0	0.0	0.7	0.0	100.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	100.0	100.0	0.0	100.0	100.0	0.0	0.0	0.0
	C	0.0	0.0	0.0	0.0	98.9	88.0	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	G	100.0	0.0	100.0	0.0	0.2	2.3	99.9	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	100.0	100.0
	T	0.0	0.0	0.0	100.0	1.0	8.9	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table S2. Base editing outcomes from treatment with BE3, CDA1-BE3, AID-BE3, or APOBEC3G-BE3 at the *FANCF* locus. The sequence of the protospacer is shown at the top, with the PAM in blue and the target bases in red with a subscripted number indicating their positions within the protospacer. Underneath the sequence are the percentages of total sequencing reads with the corresponding base. Cells were treated as described in the Methods. Values shown are from one representative experiment.

FANCF		G	G	A	A	T	C_6	C_7	C_8	T	T	C_{11}	T	G	C	A	G	C	A	C	C	T	G	G
untreated	A	0.1	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	0.0	0.0
	G	99.9	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0
	T	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0
BE3	A	0.0	0.1	100.0	100.0	0.0	1.0	1.4	1.4	0.0	0.0	0.2	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	0.0	0.0	0.0	70.7	75.0	74.4	0.0	0.0	88.8	0.0	0.0	99.9	0.0	0.0	100.0	0.0	100.0	99.9	0.0	0.0	0.0
	G	100.0	99.9	0.0	0.0	0.0	0.4	0.8	0.9	0.0	0.0	0.1	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	99.9	100.0
	T	0.0	0.0	0.0	0.0	100.0	28.0	22.9	23.4	100.0	100.0	10.9	100.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.1	100.0	0.0	0.0
CDA1-BE3	A	0.1	0.1	100.0	100.0	0.0	0.3	0.6	0.3	0.0	0.0	0.0	0.0	0.1	0.0	100.0	0.1	0.0	100.0	0.0	0.0	0.0	0.1	0.1
	C	0.0	0.0	0.0	0.0	0.0	82.4	83.5	84.7	0.0	0.0	97.3	0.0	0.0	99.6	0.0	0.0	99.8	0.0	100.0	100.0	0.0	0.0	0.0
	G	99.9	99.9	0.0	0.0	0.0	0.1	0.3	0.2	0.0	0.0	0.0	0.0	99.9	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	99.9	99.9
	T	0.0	0.0	0.0	0.0	100.0	17.2	15.6	14.7	99.9	100.0	2.6	100.0	0.0	0.3	0.0	0.0	0.2	0.0	0.0	0.0	100.0	0.0	0.0
AID-BE3	A	0.0	0.0	99.9	99.9	0.1	0.2	0.6	0.3	0.0	0.0	0.0	0.0	0.1	0.0	99.9	0.1	0.1	100.0	0.0	0.0	0.0	0.1	0.1
	C	0.0	0.0	0.0	0.0	0.0	73.8	80.3	86.6	0.0	0.0	96.7	0.0	0.0	98.8	0.0	0.0	97.0	0.0	99.9	100.0	0.0	0.0	0.0
	G	99.9	99.9	0.0	0.0	0.0	0.8	0.5	0.4	0.0	0.0	0.0	0.0	99.9	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	99.9	99.9
	T	0.0	0.0	0.0	0.0	99.9	25.2	18.6	12.7	100.0	100.0	3.2	100.0	0.0	1.2	0.0	0.0	2.9	0.0	0.1	0.0	100.0	0.0	0.0
APOBEC3G-BE3	A	0.0	0.0	100.0	100.0	0.0	0.0	2.9	1.7	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.1	0.0	100.0	0.0	0.0	0.0	0.0	0.1
	C	0.0	0.0	0.0	0.0	0.0	96.1	68.3	68.3	0.0	0.0	99.8	0.0	0.0	99.9	0.0	0.0	100.0	0.0	100.0	100.0	0.0	0.0	0.0
	G	100.0	99.9	0.0	0.0	0.0	0.3	1.6	10.8	0.0	0.0	0.0	0.0	100.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	100.0	99.9
	T	0.0	0.0	0.0	0.0	100.0	3.6	27.2	19.2	100.0	100.0	0.1	100.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0

Table S3. Base editing outcomes from treatment with BE3, CDA1-BE3, AID-BE3, or APOBEC3G-BE3 at the *HEK2* locus. The sequence of the protospacer is shown at the top, with the PAM in blue and the target bases in red with a subscripted number indicating their positions within the protospacer. Underneath the sequence are the percentages of total sequencing reads with the corresponding base. Cells were treated as described in the Methods. Values shown are from one representative experiment.

HEK2		G	A	A	C_4	A	C_6	A	A	A	G	C_{11}	A	T	A	G	A	C	T	G	C	G	G	G
untreated	A	0.0	100.0	100.0	0.0	100.0	0.0	100.0	100.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	0.0	100.0	0.0	99.9	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	99.9	0.0	0.0	0.0
	G	100.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	99.9	100.0	100.0
	T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.1	0.0	0.0	0.0
BE3	A	0.1	100.0	100.0	1.3	99.9	0.7	100.0	100.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	100.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	0.0	80.5	0.0	48.5	0.0	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	99.7	0.0	0.0	100.0	0.0	0.0	0.0
	G	99.9	0.0	0.0	0.9	0.1	36.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	100.0	99.9	100.0
	T	0.0	0.0	0.0	17.2	0.0	14.8	0.0	0.0	0.0	0.0	0.1	0.0	100.0	0.0	0.0	0.0	0.2	100.0	0.0	0.0	0.0	0.0	0.0
CDA1-BE3	A	0.0	99.9	99.9	9.6	99.9	1.8	100.0	100.0	100.0	0.0	0.2	100.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	0.0	70.3	0.0	73.5	0.0	0.0	0.0	0.0	90.7	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	99.9	0.0	0.0	0.0
	G	100.0	0.0	0.1	0.3	0.1	3.5	0.0	0.0	0.0	100.0	0.3	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	100.0
	T	0.0	0.0	0.0	19.8	0.0	21.2	0.0	0.0	0.0	0.0	8.8	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.1	0.0	0.0	0.0
AID-BE3	A	0.0	99.9	99.9	6.9	99.9	1.2	100.0	100.0	100.0	0.0	0.2	100.0	0.0	100.0	0.0	100.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	0.0	54.5	0.0	52.4	0.0	0.0	0.0	0.0	83.8	0.0	0.0	0.0	0.0	0.0	99.7	0.0	0.0	99.8	0.0	0.0	0.0
	G	99.9	0.0	0.1	0.7	0.1	6.1	0.0	0.0	0.0	99.9	0.6	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	100.0
	T	0.1	0.1	0.0	37.9	0.0	40.3	0.0	0.0	0.0	0.1	15.5	0.0	100.0	0.0	0.0	0.0	0.2	99.9	0.0	0.2	0.0	0.0	0.0
APOBEC3G-BE3	A	0.0	100.0	100.0	0.1	100.0	0.1	100.0	100.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	0.0	97.8	0.0	80.3	0.0	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	99.9	0.0	0.0	0.0
	G	100.0	0.0	0.0	0.1	0.0	5.9	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	100.0
	T	0.0	0.0	0.0	2.0	0.0	13.7	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.1	0.0	0.0	0.0

Table S4. Base editing outcomes from treatment with BE3, CDA1-BE3, AID-BE3, or APOBEC3G-BE3 at the *HEK3* locus. The sequence of the protospacer is shown at the top, with the PAM in blue and the target bases in red with a subscripted number indicating their positions within the protospacer. Underneath the sequence are the percentages of total sequencing reads with the corresponding base. Cells were treated as described in the Methods. Values shown are from one representative experiment.

HEK3		G	G	C_3	C_4	C_5	A	G	A	C_9	T	G	A	G	C	A	C	G	T	G	A	T	G	G
untreated	A	0.0	0.1	0.0	0.0	0.0	99.9	0.0	100.0	0.0	0.0	0.0	99.9	0.0	0.0	99.9	0.0	0.0	0.0	0.0	99.9	0.0	0.0	0.1
	C	0.0	0.0	100.0	99.9	99.9	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0
	G	100.0	99.9	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	99.9	0.0	99.9	0.0	0.0	0.0	99.9	0.0	100.0	0.0	0.0	100.0	99.9
	T	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0
BE3	A	0.0	0.0	0.0	0.8	1.1	100.0	0.0	100.0	0.4	0.0	0.0	100.0	0.2	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0
	C	0.0	0.0	98.6	67.4	58.1	0.0	0.0	0.0	98.3	0.0	0.0	0.0	0.0	100.0	0.0	99.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	G	100.0	100.0	0.0	0.4	8.9	0.0	100.0	0.0	0.6	0.0	100.0	0.0	99.8	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	99.9	100.0
	T	0.0	0.0	1.3	31.4	31.9	0.0	0.0	0.0	0.7	100.0	0.0	0.0	0.0	0.0	0.0	0.3	0.0	100.0	0.0	0.0	100.0	0.1	0.0
CDA1-BE3	A	0.1	0.0	0.8	0.6	0.0	100.0	0.0	99.9	0.4	0.0	0.0	99.9	0.0	0.0	100.0	0.0	0.0	0.0	0.0	99.9	0.0	0.1	0.1
	C	0.0	0.0	79.4	73.6	73.4	0.0	0.0	0.0	91.0	0.0	0.0	0.0	0.0	99.9	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	G	99.9	100.0	0.0	0.0	2.1	0.0	100.0	0.0	0.3	0.0	99.9	0.0	100.0	0.0	0.0	0.0	99.9	0.0	99.9	0.1	0.0	99.9	99.9
	T	0.1	0.0	19.8	25.9	24.4	0.0	0.0	0.0	8.3	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	0.0	100.0	0.0	0.0
AID-BE3	A	0.1	0.3	1.0	0.7	0.3	99.9	0.0	100.0	1.6	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0
	C	0.0	0.0	63.9	75.4	60.2	0.0	0.0	0.0	86.6	0.0	0.0	0.0	0.0	98.9	0.0	99.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	G	99.8	99.7	0.1	0.4	1.0	0.0	100.0	0.0	1.2	0.0	100.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	100.0	99.9
	T	0.0	0.0	35.0	23.5	38.5	0.0	0.0	0.0	10.6	100.0	0.0	0.0	0.0	1.1	0.0	0.2	0.0	100.0	0.0	0.0	100.0	0.0	0.0
APOBEC3G-BE3	A	0.0	0.0	0.0	0.6	3.1	100.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0
	C	0.0	0.0	99.9	86.7	33.9	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	G	100.0	99.9	0.0	1.4	28.6	0.0	100.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	100.0	100.0
	T	0.0	0.0	0.0	11.3	34.4	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0

Table S5. Base editing outcomes from treatment with BE3, CDA1-BE3, AID-BE3, or APOBEC3G-BE3 at the *HEK4* locus. The sequence of the protospacer is shown at the top, with the PAM in blue and the target bases in red with a subscripted number indicating their positions within the protospacer. Underneath the sequence are the percentages of total sequencing reads with the corresponding base. Cells were treated as described in the Methods. Values shown are from one representative experiment.

HEK4		G	G	C_3	A	C_5	T	G	C_8	G	G	C	T	G	G	A	G	G	T	G	G	G	G	G
untreated	A	0.0	0.1	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	100.0	0.0	100.0	0.0	0.0	99.9	0.0	0.0	100.0	0.0	0.0	0.0	0.1	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0
	G	99.9	99.9	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	0.0	100.0	100.0	0.0	100.0	100.0	0.4	99.9	99.9	100.0	99.9	99.9
	T	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	99.5	0.1	0.0	0.0	0.0	0.0
BE3	A	0.1	0.0	0.1	99.9	8.4	0.1	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.0	99.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	99.0	0.0	41.9	0.0	0.0	99.3	0.0	0.0	99.8	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	G	99.9	99.9	0.0	0.0	24.5	0.0	100.0	0.0	100.0	99.9	0.0	0.0	100.0	100.0	0.1	100.0	100.0	0.0	100.0	100.0	100.0	100.0	100.0
	T	0.0	0.0	0.9	0.0	25.2	99.9	0.0	0.5	0.0	0.0	0.2	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0
CDA1-BE3	A	0.2	0.1	0.5	100.0	0.5	0.0	0.0	0.2	0.0	0.0	0.1	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	91.3	0.0	87.5	0.0	0.0	93.4	0.0	0.0	97.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	G	99.8	99.9	0.3	0.0	0.0	0.0	100.0	0.0	100.0	100.0	0.0	0.0	100.0	100.0	0.0	100.0	100.0	0.0	100.0	100.0	100.0	100.0	100.0
	T	0.0	0.0	7.8	0.0	12.0	100.0	0.0	6.3	0.0	0.0	2.9	100.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0
AID-BE3	A	0.3	0.4	0.3	100.0	0.9	0.0	0.0	0.9	0.0	0.0	0.1	0.0	0.0	0.0	99.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	81.5	0.0	71.3	0.0	0.0	80.7	0.0	0.0	97.9	0.1	0.0	0.0	0.2	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0
	G	99.7	99.6	0.1	0.0	0.7	0.0	100.0	0.1	100.0	100.0	0.0	0.0	99.9	100.0	0.0	100.0	100.0	0.2	99.9	100.0	100.0	100.0	99.9
	T	0.0	0.0	18.0	0.0	27.1	100.0	0.0	18.3	0.0	0.0	2.0	99.9	0.0	0.0	0.0	0.0	0.0	99.8	0.0	0.0	0.0	0.0	0.1
APOBEC3G-BE3	A	0.1	0.0	0.0	99.9	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	C	0.0	0.0	99.9	0.0	99.4	0.0	0.0	99.9	0.0	0.0	100.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0
	G	99.9	99.9	0.0	0.0	0.4	0.0	99.9	0.0	100.0	99.9	0.0	0.0	99.9	100.0	0.1	100.0	99.9	0.0	99.9	100.0	100.0	100.0	99.9
	T	0.0	0.0	0.0	0.0	0.2	99.9	0.0	0.1	0.0	0.1	0.0	99.9	0.1	0.0	0.0	0.0	0.1	99.9	0.0	0.0	0.0	0.0	0.1

Table S6. Base editing outcomes from treatment with BE3, CDA1-BE3, AID-BE3, or APOBEC3G-BE3 at the *RNF2* locus. The sequence of the protospacer is shown at the top, with the PAM in blue and the target bases in red with a subscripted number indicating their positions within the protospacer. Underneath the sequence are the percentages of total sequencing reads with the corresponding base. Cells were treated as described in the Methods. Values shown are from one representative experiment.

RNF2		G	T	C_3	A	T	C_6	T	T	A	G	T	C_{12}	G	T	T	A	C	C	T	G	A	G	G
untreated	A	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0
	C	0.0	0.0	99.9	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0
	G	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0
	T	0.0	100.0	0.1	0.0	100.0	0.0	100.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0
BE3	A	0.0	0.0	0.3	99.9	0.1	1.6	0.1	0.0	100.0	0.0	0.0	0.1	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0
	C	0.0	0.0	81.1	0.0	0.0	59.7	0.0	0.0	0.0	0.0	0.0	96.6	0.0	0.0	0.0	0.0	99.9	99.9	0.0	0.0	0.0	0.0	0.0
	G	99.9	0.0	0.0	0.0	0.0	12.6	0.0	0.0	0.0	100.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	99.9	100.0
	T	0.0	100.0	18.6	0.0	99.9	26.1	99.9	100.0	0.0	0.0	100.0	3.3	0.0	100.0	100.0	0.0	0.0	0.1	100.0	0.0	0.0	0.0	0.0
CDA1-BE3	A	0.0	0.0	0.3	100.0	0.0	0.3	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0
	C	0.0	0.0	88.1	0.0	0.0	87.8	0.0	0.0	0.0	0.0	0.0	99.1	0.0	0.0	0.0	0.0	99.9	99.9	0.0	0.0	0.0	0.0	0.0
	G	100.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	100.0
	T	0.0	100.0	11.6	0.0	100.0	11.5	100.0	100.0	0.0	0.0	100.0	0.9	0.0	100.0	100.0	0.0	0.1	0.1	100.0	0.0	0.0	0.0	0.0
AID-BE3	A	0.0	0.0	0.1	100.0	0.1	0.1	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0
	C	0.0	0.0	92.7	0.0	0.0	81.5	0.0	0.0	0.0	0.0	0.0	99.8	0.0	0.0	0.0	0.0	98.7	99.8	0.0	0.0	0.0	0.0	0.0
	G	100.0	0.0	0.0	0.0	0.0	3.2	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	99.9	0.0	99.9	100.0
	T	0.0	100.0	7.2	0.0	99.9	15.2	100.0	100.0	0.0	0.0	100.0	0.1	0.0	100.0	100.0	0.0	1.2	0.2	100.0	0.0	0.0	0.0	0.0
APOBEC3G-BE3	A	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0
	C	0.0	0.0	98.2	0.0	0.0	82.8	0.0	0.0	0.0	0.0	0.0	99.8	0.0	0.0	0.0	0.0	99.8	99.8	0.0	0.0	0.0	0.0	0.0
	G	100.0	0.0	0.0	0.0	0.0	3.4	0.0	0.0	0.0	99.9	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	100.0	100.0
	T	0.0	100.0	1.8	0.0	100.0	13.8	100.0	100.0	0.0	0.0	100.0	0.2	0.0	100.0	100.0	0.0	0.2	0.2	100.0	0.0	0.0	0.0	0.0

SUPPLEMENTARY NOTES

Note S1. Python script to detect linkage disequilibrium in base editing outcomes at targets sites with multiple target cytidines. An example script is shown here for a single target site (RNF2). Multiple variables were altered for each site including: indir, outdir, 7 nucleotide sequences used to define the protospacer as well as the position within the protospacer with the highest frequency of base editing byproducts.

```
%matplotlib inline
import numpy as np
import scipy as sp
import matplotlib as mpl
import matplotlib.cm as cm
import matplotlib.pyplot as plt
import pandas as pd
pd.set_option('display.width', 500)
pd.set_option('display.max_columns', 100)
pd.set_option('display.notebook_repr_html', True)
import seaborn as sns
sns.set_style("whitegrid")
sns.set_context("poster")
import requests
import time from bs4
import BeautifulSoup
import regex
import re
import os
from Bio import SeqIO
import Bio
from Bio import motifs
#BE processivity analysis, RNF2
site='_RNF2'
indir='/Users/michaelpacker/Desktop/Liu_Lab/MiSeqData/ACK060717/fastq/RNF2/'
outdir='/Users/michaelpacker/Desktop/Liu_Lab/MiSeqData/ACK060717/fastq/RNF2/'
filenames=os.listdir(indir)
for i in range(len(filenames)):
     seqs={}
     if filenames[i][-5:]=='fastq':
         for record in SeqIO.parse(indir+filenames[i], "fastq") :
             #split prior to spacer window
             split1=record.seq.tostring().split('CTTGGCA')
             if len(split1)==2:
                 #take second item in first split, and split again at the seq
uence right after the protospacer and take first item
                 split2=split1[1].split('AGGTGTT')[0]
                 #keep only 20 basepair long protospacers
                 if (len(split2)==20) & (split2.find('N')==-1):
                     seqs[record.id]=split2
         #generate dataframe with protospacer sequence column and identity of
the nucleotide at the position within the protospacer with the highest freque
ncy of editing byproducts ("DirtyPosition")
         frame=pd.DataFrame({'Spacer':seqs.values(), 'DirtyPosition':[x[5] fo
r x in seqs.values()]}, index=seqs.keys())
         #generate a biopython motif for all the protospacer sequences
         MotifAll=motifs.create(frame.Spacer.values)
         #generate a biopython motif for all protospacer sequences with (A,C,
G,T) at the DirtyPosition if no sequences have an (A,C,G,T) at the DirtyPosit
ion save a placeholder motif with a single polyA sequence
         if len(frame[frame.DirtyPosition=='A'])>0:
```

```
            MotifA=motifs.create(frame[frame.DirtyPosition=='A'].Spacer.valu
es)
        else:
            MotifA=motifs.create(['A'*20])
        if len(frame[frame.DirtyPosition=='C'])>0:
            MotifC=motifs.create(frame[frame.DirtyPosition=='C'].Spacer.valu
es)
        else:
            MotifC=motifs.create(['A'*20])
        if len(frame[frame.DirtyPosition=='G'])>0:
            MotifG=motifs.create(frame[frame.DirtyPosition=='G'].Spacer.valu
es)
        else:
            MotifG=motifs.create(['A'*20])
        if len(frame[frame.DirtyPosition=='T'])>0:
            MotifT=motifs.create(frame[frame.DirtyPosition=='T'].Spacer.valu
es)
        else:
            MotifT=motifs.create(['A'*20])
        #Save motif counts in 10 DataFrames for the 4 groups and the total p
ool of sequences (both raw and normalized for each)
        a=pd.DataFrame(MotifA.counts, index=['A'+str(s) for s in range(20)])
        A=pd.DataFrame(MotifA.counts.normalize(),index=['A'+str(s) for s in
range(20)])
        c=pd.DataFrame(MotifC.counts, index=['C'+str(s) for s in range(20)])
        C=pd.DataFrame(MotifC.counts.normalize(), index=['C'+str(s) for s in
range(20)])
        g=pd.DataFrame(MotifG.counts, index=['G'+str(s) for s in range(20)])
        G=pd.DataFrame(MotifG.counts.normalize(), index=['G'+str(s) for s in
range(20)])
        t=pd.DataFrame(MotifT.counts, index=['T'+str(s) for s in range(20)])
        T=pd.DataFrame(MotifT.counts.normalize(), index=['T'+str(s) for s in
range(20)])
        All=pd.DataFrame(MotifAll.counts, index=['All'+str(s) for s in range
(20)])
        ALL=pd.DataFrame(MotifAll.counts.normalize(),index=['All'+str(s) for
s in range(20)])
        #export csv files for the motif counts (both raw and normalized)
        All.append(a).append(c).append(g).append(t).to_csv(outdir+filenames[
i].strip('.fastq')+'RawMotifs.csv')
        ALL.append(A).append(C).append(G).append(T).to_csv(outdir+filenames[
i].strip('.fastq')+'NormalizedMotifs.csv')
        #export csv file with abundance of each unique protospacer sequence
        Counts=pd.DataFrame(seqs.items(), columns=['ID','Window']).groupby('
Window').count().sort('ID', ascending=False)
        Counts.to_csv(outdir+filenames[i].strip('.fastq')+'.csv')
```

SUPPLEMENTARY SEQUENCES

Amino Acid Sequences of CDA1-BE3, AID-BE3, BE3-Gam, SaBE3-Gam, BE4, BE4-Gam, SaBE4, and SaBE4-Gam

CDA1-BE3:
MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQSGTERGIHAEIF
SIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHTLKIWACKLYYEKNARNQIGLW
NLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEKRRSELSIMIQVKILHTTKSPAVSG
SETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHE
KYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINA
SGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDL
DNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYK
EIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFI
ERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVK
QLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL
KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKE
DIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKN
SRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFL
KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFI
KRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYL
NAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPL
IETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYG
GFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF
SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH
QSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDEST
DENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV

AID-BE3:
MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELLFLRYISDWD
LDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQIAI
MTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGLSGSETPGTS
ESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR
TARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHL
RKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAI
LSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIG
DQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSK
NGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQED
FYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFD
KNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFK
KIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD
DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVS
GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRI
EEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNK
VLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR
QITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTAL
IKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG
EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVA
YSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRM
LASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADA
NLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYET
RIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTS
DAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV

BE3-Gam:

MAKPAKRIKSAAAAYVPQNRDAVITDIKRIGDLQREASRLETEMNDAIAEITEKFAARIAPIKTDIETLSKGVQ
GWCEANRDELTNGGKVVKTANLVTGDVSWRVRPPSVSIRGMDAVMETLERLGLQRFIRTKQEINKEAILLE
PKAVAGVAGITVKSGIEDFSIIPFEQEAGISGSETPGTSESATPESSSETGPVAVDPTLRRRIEPHEFEVFF
DPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECS
RAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWP
RYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPGTSES
ATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA
RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLR
KKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAIL
SARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGD
QYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN
GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDF
YPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDK
NLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKK
IECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDD
KVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSG
QGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEE
GIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL
TRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI
TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIK
KYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEI
VWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAY
SVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRML
ASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADAN
LDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI
DLSQLGGDSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSD
APEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV

SaBE3-Gam:
MAKPAKRIKSAAAAYVPQNRDAVITDIKRIGDLQREASRLETEMNDAIAEITEKFAARIAPIKTDIETLSKGVQ
GWCEANRDELTNGGKVVKTANLVTGDVSWRVRPPSVSIRGMDAVMETLERLGLQRFIRTKQEINKEAILLE
PKAVAGVAGITVKSGIEDFSIIPFEQEAGISGSETPGTSESATPESSSETGPVAVDPTLRRRIEPHEFEVFF
DPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECS
RAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWP
RYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPGTSES
ATPESGKRNYILGLAIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQR
VKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKE
QISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETR
RTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEY
YEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIA
KILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKK
VDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQT
NERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLV
KQEENSKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLV
DTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKE
WKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTL
YSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYY
EETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVK
NLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYRE
YLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKGGSPKKKRKVSSDYKDHDGDYKD
HDIDYKDDDDKSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLT
SDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV

BE4:
MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKF
TTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQ
IMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSC

HYQRLPPHILWATGLKSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVIT
DEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV
DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFR
GHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGL
FGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT
EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEK
MDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPL
ARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELT
KVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHD
LLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLING
IRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTV
KVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKL
YLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMK
NYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIR
EVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMI
AKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNI
VKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELL
GITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFL
YLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAEN
IIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSGGSGGSTNL
SDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNG
ENKIKMLSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVM
LLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRK

BE4-Gam:
MAKPAKRIKSAAAAYVPQNRDAVITDIKRIGDLQREASRLETEMNDAIAEITEKFAARIAPIKTDIETLSKGVQ
GWCEANRDELTNGGKVKTANLVTGDVSWRVRPPSVSIRGMDAVMETLERLGLQRFIRTKQEINKEAILLE
PKAVAGVAGITVKSGIEDFSIIPFEQEAGISGSETPGTSESATPESSSETGPVAVDPTLRRRIEPHEFEVFF
DPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECS
RAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWP
RYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGGSSGGSSGS
ETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGAL
LFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNI
VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQ
LFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS
KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVR
QQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIP
HQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVD
KGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFK
TNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFED
REMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIH
DDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQT
TQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVD
HIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLS
ELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNY
HHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLA
NGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKK
DWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLII
KLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYL
DEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTK
EVLDATLIHQSITGLYETRIDLSQLGGDSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNK
PESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGSTNLSDIIEKETGK
QLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSG
GSPKKKRK

SaBE4:
MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKF
TTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQ

IMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSC
HYQRLPPHILWATGLKSGGSSGGSSGSETPGTSESATPESSGGSSGGSGKRNYILGLAIGITSVGYGIIDY
ETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARV
KGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGE
VRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEML
MGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILV
NEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEI
EQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSF
IQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQ
EGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISYE
TFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVK
SINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEI
ETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKL
KKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKL
NAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISN
QAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTD
ILGNLYEVKSKKHPQIIKKGGSPKKKRKVSSDYKDHDGDYKDHDIDYKDDDDKSGGSGGSGGSTNLSDII
EKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGEN
KIKMLSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLL
TSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV

SaBE4-Gam:
MAKPAKRIKSAAAAYVPQNRDAVITDIKRIGDLQREASRLETEMNDAIAEITEKFAARIAPIKTDIETLSKGVQ
GWCEANRDELTNGGKVKTANLVTGDVSWRVRPPSVSIRGMDAVMETLERLGLQRFIRTKQEINKEAILLE
PKAVAGVAGITVKSGIEDFSIIPFEQEAGISGSETPGTSESATPESSSETGPVAVDPTLRRRIEPHEFEVFF
DPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECS
RAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWP
RYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGGSSGGSSGS
ETPGTSESATPESSGGSSGGSGKRNYILGLAIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSK
RGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHN
VNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAY
HQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALN
DLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIK
DITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWH
TNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSK
DAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHI
IPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEER
DINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKH
HAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSH
RVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIM
EQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFD
VYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDL
LNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKGGSPKKKRKV
SSDYKDHDGDYKDHDIDYKDDDDKSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPE
SDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGSTNLSDIIEKETGKQL
VIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGS
PKKKRKV