

NUCLEOBASE EDITORS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of priority of Singapore provisional application No. 10201913340Q, filed on 26 December 2019, the contents of it being hereby incorporated by reference in its entirety for all purposes.

FIELD OF THE INVENTION

[0002] The present invention relates to the field of molecular biology biotechnology, specifically the field of gene editing and more specifically nucleobase editing.

BACKGROUND OF THE INVENTION

[0003] Many human genetic diseases are caused by single-nucleotide polymorphisms (SNPs), in which the disease and healthy alleles differ by a single DNA base. CRISPR-Cas9 nuclease is commonly used to edit genomic DNA in a targetable fashion. Following DNA cleavage, three major repair mechanisms can be involved in fixing that break – homology directed repair (HDR), micro-homology mediated end joining (MMEJ), and non-homologous end joining (NHEJ). In the presence of a donor DNA/RNA, HDR may occur. However, previous studies were only able to achieve low levels of precise gene editing (0.1 to 5%). MMEJ requires that the double stranded break be formed at a region with micro-homology and hence restricts the targeting range of CRISPR-Cas9. NHEJ is the predominant pathway in repairing Cas9-induced double stranded breaks. Unfortunately, it introduces a variety of random indels. For therapeutic applications where precise point mutations are necessary, NHEJ is unable to restore a defective gene and is hence inadequate.

[0004] Base editors can correct these SNPs by converting the targeted DNA bases into another base in a controllable and efficient fashion. Current technology enables the conversion of C•G base pairs to T•A base pairs using cytosine base editors (CBEs) (*A. C. Komor, et al., Nature 533, 420-424 (2016); K. Nishida et al., Science 353, aaf8729 (2016); A. C. Komor, et al., Sci. Adv. 3, eaao4774 (2017)*) and A•T base pairs to G•C base pairs using adenine base editors (ABEs) (*N. M. Gaudelli, et al., Nature 551, 464-471 (2017)*), which together represent half of all known disease-associated SNPs. CBEs and ABEs are also known to effect some C•G to G•C edits as byproducts, but they cannot effect such transversions at efficiencies or purities necessary for therapeutic use, and hence the remaining half of these SNPs are not addressable by current base-editors.

[0005] Therefore, there is a need to provide novel nucleobase editors which facilitate C:G to G:C editing (CGBEs) with high efficiency, specificity, and purity.

SUMMARY OF THE INVENTION

[0006] In one aspect, the present disclosure refers to a fusion protein or a protein complex comprising a DNA binding protein (DnaBP), a nucleobase modifying protein (NMP), and a Base Excision Repair associated protein (BERAP); wherein the fusion protein or protein complex does not comprise a Uracil binding protein or a catalytically active DNA polymerase.

[0007] In another aspect, the present disclosure refers to a fusion protein comprising:

- a first amino acid sequence that is at least 80% identical to the amino acid sequence of any one of SEQ ID NOs: 1-2,
- a second amino acid sequence that is at least 80% identical to the amino acid sequence of SEQ ID NO: 3,
- a third amino acid sequence that is at least 80% identical to the amino acid sequence of SEQ ID NO: 4-10.

[0008] In another aspect, the present disclosure refers to a fusion protein comprising a sequence of any one of SEQ ID NOs: 42 to 72.

[0009] In yet another aspect, the present disclosure refers to protein complex comprising:

- a first protein comprising an amino acid sequence that is at least 80% identical to the amino acid sequence of any one of SEQ ID NOs: 1-2,
- a second protein comprising an amino acid sequence that is at least 80% identical to the amino acid sequence of SEQ ID NO: 3,
- a third protein comprising an amino acid sequence that is at least 80% identical to the amino acid sequence of SEQ ID NO: 4-11.

[0010] In yet another aspect, the present disclosure refers to a protein-nucleic acid complex comprising a nucleic acid molecule and any one of:

- the fusion protein or protein complex as disclosed herein,
- the fusion protein as disclosed herein; and
- the protein complex of claim 20.

[0011] In yet another aspect, the present disclosure refers to a pharmaceutical composition comprising the fusion protein or protein complex as disclosed herein, the fusion protein as disclosed herein, the protein complex as disclosed herein, or the protein-nucleic acid complex as disclosed herein.

[0012] In yet another aspect, the present disclosure refers to a method of replacing a cytosine with a guanine on a DNA strand in a cell, said method comprises introducing to the cell the fusion protein or protein complex as disclosed herein, the fusion protein as disclosed herein, the protein complex as disclosed herein, or the protein-nucleic acid complex as disclosed herein.

[0013] In yet another aspect, the present disclosure refers to a polynucleotide encoding the fusion protein or protein complex as disclosed herein, the fusion protein as disclosed herein, the protein complex as disclosed herein, or the protein-nucleic acid complex as disclosed herein.

[0014] In yet another aspect, the present disclosure refers to a vector comprising the polynucleotide as disclosed herein.

[0015] In yet another aspect, the present disclosure refers to a cell comprising the fusion protein or protein complex as disclosed herein, the fusion protein as disclosed herein, the protein complex as disclosed herein, or the protein-nucleic acid complex as disclosed herein.

[0016] In yet another aspect, the present disclosure refers to a method of treating a subject having or suspected of having a disease or disorder comprising administering the fusion protein or protein complex as disclosed herein, the fusion protein as disclosed herein, the protein complex as disclosed herein, the protein-nucleic acid complex as disclosed herein, the pharmaceutical composition as disclosed herein, the polynucleotide as disclosed herein, or the vector as disclosed herein to the subject.

[0017] In yet another aspect, the present disclosure refers to a method for editing a target nucleobase pair of a double-stranded DNA sequence, the method comprising:

- a. contacting a target region of the double-stranded DNA sequence with a complex comprising a nucleobase editor and a guide nucleic acid, wherein the target region comprises the target nucleobase pair;

- b. inducing strand separation of said target region;

- c. converting a first nucleobase of said target nucleobase pair in a single strand of the target region to a second nucleobase;

- d. excising said second nucleobase from the double-stranded DNA sequence to produce an abasic site; and

- e. promoting the Base Excision Repair (BER) pathway to repair the abasic site, generating a third nucleobase at the abasic site, wherein the third nucleobase is different from the first nucleobase.

BRIEF DESCRIPTION OF THE DRAWINGS

[0018] The invention will be better understood with reference to the detailed description when considered in conjunction with the non-limiting examples and the accompanying drawings, in which:

[0019] **Figure 1** shows 2 illustrations: one showing the difference between C:G to G:C Base Editors (CGBE) and (Cytosine base editors) CBE; the other showing the CGBE candidates; and 1 column graph. (a) CBEs like BE3 and BE4 predominantly convert C:G to T:A while CGBE aims to predominantly convert C:G to G:C. (b) CGBE candidates were designed in three orientations – ACX, AXC, and XAC, where X denotes the fused BER protein (or Base Excision Repair associated protein, BERAP). (c) Seven candidates were selected for their high C:G to G:C editing at both HEK2 and HEK3. Targeted C's are denoted in a box. PAMs are underlined. The lower editing at HEK3 is likely due to a disfavored motif (refer to data in Fig. 2a). * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (one-way ANOVA against 'Untreated,' Dunn-Šidák; $n=3$, mean $\pm$ standard error). Figure 1 describes an initial screen of CGBE candidates for C:G to G:C editing.

[0020] **Figure 2** shows 2 column graphs and 1 image showing the quantitative representation of the C:G to G:C editing. (a) Evaluation of C:G to G:C editing at each NCN DNA motif. 16 different gRNAs were designed to target the genomic region around the HEK2 site (HEK2-1 to HEK2-16 in Extended Data Table 2), chosen such that the gRNA-to-target combinations together cover all NCN motif contexts and that genomic distance among gRNAs are minimized (14/16 gRNAs, including the initial HEK2-1 gRNA:target, reside within a 1.8kb region, while the other 2 gRNAs target within 10kb). ($n=2$, mean $\pm$ standard error). (b) DNA WebLogo created with target motifs in which C:G at position 6 was edited to G:C ($n=2$; error bars are Bayesian 95% confidence intervals). (c) Editing window of CGBEs using gRNAs with alternating 5'-W-C-3' motifs. Targeted C's are in a larger font. PAMs are underlined. * $p < 0.05$; ** $p < 0.01$ (one-way ANOVA against 'BE3,' Dunn-Šidák; $n=3$, mean $\pm$ standard error). Figure 2 describes the sequence context and editing window for two selected CGBEs.

[0021] **Figure 3** shows 2 scatter plots. (a) Removal of UGI from BE3 increases C:G to G:C editing (triangles); fusion of rXRCC1 further increases C:G to G:C editing. The major byproduct is C:G to T:A editing (circles). (b) Mean C:G to G:C editing/C:G to T:A editing ratio. Data includes all biological replicates across 16 genomic loci that reside within a WCW, ACC or GCT motif. * $p < 0.05$; *** $p < 0.001$ (two-tailed Student's T test; $n=28-45$ biological replicates). Figure 3 illustrates how CGBE induces efficient C:G to G:C editing as the predominant product.

[0022] **Figure 4** shows 2 column graphs. (a) For some CGBE candidates, C:G to G:C editing is the predominant edit at position 6. (b) C:G to T:A editing is the predominant edit at

position 4. The seven candidates selected for further studies are marked with †. Targeted C's are in a larger font. PAMs are underlined. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (one-way ANOVA against 'Untreated,' Dunn-Sidak; $n=3$, mean $\pm$ standard error). Figure 4 shows the CGBE Initial screen of CGBE candidates on HEK2-1.

[0023] Figure 5 shows 3 column graphs representing the initial screen of CGBE candidates on HEK3 at (a) position 5, (b) position 4, and (c) position 3. The seven candidates selected for further studies are marked with †. Targeted C is in a larger font. PAM is underlined. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (one-way ANOVA against 'Untreated,' Dunn-Sidak; $n=3$, mean $\pm$ standard error). Figure 5 illustrates the results of the initial screen of CGBE candidates on HEK3.

[0024] Figure 6 shows 3 column graphs. (a) CGBE candidates effect C:G to G:C mutations at EMX1, HEK4, RNF2, and FANCF. C:G to G:C editing is the main edit at HEK4 and RNF2; C:G to T:A editing is the main edit at FANCF and EMX1. (one-way ANOVA against 'BE3,' Dunn-Sidak; $n=3$, mean $\pm$ standard error) (b) CGBE editing at disease-associated genes ADRB2, GJB2, MYBPC3, and GAL 292. Note that ADRB2 contains naturally occurring polymorphism in HEK293AAV cells, and hence this data is not included in Fig 3. (one-way ANOVA against 'Untreated,' Dunn-Sidak; $n=5$ for ADRB2 and MYBPC3; $n=4$ for GJB2; $n=1$ for GAL 292, mean $\pm$ standard error) (c) Mean C:G to G:C editing as a percent of all reads across 16 NCN sites. CGBEs increase C:G to G:C editing by three to four fold compared to BE3, across all possible NCN sequences. gRNA sequences are included in Extended Data Table 2. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (one-way ANOVA against 'BE3,' Dunn-Sidak; $n=32$, mean $\pm$ standard error). Figure 6 illustrates the further characterization of shortlisted CGBE candidates.

[0025] Figure 7 shows 2 column graphs. (a) CGBE candidates generate higher indel rates than BE3. ACX, rXRCC1 has the lowest indel rate of CGBE candidates. ($n=3$, mean $\pm$ standard error) (b) Removing UGI from BE3 increases indel rate; fusing BER proteins rPB(8kD) or rXRCC1 decreases indel rate modestly. ($n=3$, mean $\pm$ standard error) An additional set of gRNA:targets used here validates the conclusion from (a). While further mechanistic studies would be necessary, a possible hypothesis is that recruitment of the BER complex repairs abasic sites and the shortened persistence of these abasic sites may then lead to a lower propensity for indels. Figure 7 illustrates the indel rates of shortlisted CGBE candidates at genomic sites.

[0026] Figure 8 shows 2 images showing the representative data of CGBE editing at (a) ADRB2 and (b) MYBPC3. WT denotes wild-type untreated cells; XRCC denotes ACX,

XRCC1; rPB denotes ACX, rPB(8kD). Note that ADRB2 contains naturally occurring polymorphism in HEK293AAV cells.

[0027] **Figure 9** shows 2 scatter plots. (a) C:G to G:C editing (blue triangles) vs. C:G to T:A editing (orange circles) as percent of all reads across gRNAs used in this study. All biological replicates are included except those targeting the 10 suboptimal C:G to G:C base editing motifs (Fig. 2a and Fig. 2b) and ADRB2 due to naturally occurring polymorphism (Extended Data Fig. 3b). (b) Ratio of C:G to G:C editing to C:G to T:A editing across gRNAs used in this study. Only BE3 (no UGI) and ACX, rXRCC1 give a significantly higher ratio of C:G to G:C editing/C:G to T:A editing. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (one-way ANOVA against 'BE3,' Dunn-Šidák, $n=20-45$ biological replicates). Figure 9 illustrates that ACX, rXRCC1 is one of the best performers out of shortlisted CGBE candidates.

[0028] **Figure 10** shows 5 column graphs that illustrates the CGBE and BE3 off target activity at identified off-target sites with (a) HEK2-1 gRNA; (b) HEK3 gRNA; (c) HEK4 gRNA; (d) EMX1; and (e) FANCF. A total of 29 identified off-target sites with 68 editable C's using 5 gRNAs were tested. HEK3 and EMX1 off-target sites are Cas9 off-target sites identified via GUIDE-Seq19; HEK2, HEK4, and FANCF off-target sites are BE3 (no UGI) off-target sites identified via Digenome-Seq14. CGBE and BE3 induced $>0.1\%$ C:G to D:H edits at the same 15 off-target sites. At 2 out of these 15 positions, CGBE induced greater off-target editing frequency than BE3; at the remaining 13 sites, CGBE induced lower off-target editing frequency. 'OT5' indicates off-target 5; 'C4' indicates a 'C' at position 4. ($n=2$, mean $\pm$ standard error).

[0029] **Figure 11** shows 4 column graphs that illustrates the comparison of the CGBE of the present disclosure to (a) PE3 described in Anzalone *et al.*, and CGBEs described in Liu and Koblan at (b) HEK2, (c) FANCF, and (d) RNF2. (a) As positive controls for prime editing, we used previously published pegRNA (Addgene # 132778) targeting HEK3 and observed efficient prime editing (data not shown). For C:G to G:C editing, PE3 is as efficient as CGBE (ACX, rXRCC1) and induced lower levels of undesired edits at HEK4-1. At HEK2-1 and RNF2-1, PE3 is substantially less efficient than CGBE at HEK2 and RNF2. The results indicate that while PE3 may be able to perform C:G to G:C transversions at some sites, CGBE is a valuable tool to expand the editing capabilities of current technologies ($n=2$, mean $\pm$ standard error). For parts b, c, and d, since datasets were generated independently and in different cell types, comparisons should be made only against BE3 common to the two studies. Fusion of base excision enzymes such as UDG and UdgX decreases C:G to G:C editing beyond BE3 at 2 of 3 sites ($n=1$; Liu and Koblan, 2018). Fusion of base excision repair enzymes such as rXRCC1 increases C:G to G:C editing beyond BE3. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (one-way ANOVA against 'BE3,' Dunn-Šidák; $n=3$, mean $\pm$ standard error; this study).

[0030] **Figure 12** is a diagrammatic representation of the distinct strategies for CGBE design. This study employs a CGBE design strategy (left) where Cas9 is fused to protein(s) involved in repairing uracil-containing or abasic sites (AP). The activities of these BER proteins are expected to convert the AP to G before the nucleotide on the opposite strand is converted from G to C. In contrast, the polymerase strategy employed by Liu and Koblan (right) seeks to maintain the abasic site throughout a translesion synthesis envisioned to occur on the opposite strand. The AP is repaired after the nucleotide on the opposite strand is converted from G to C. The UNG-based CGBE strategy employed by Kurt *et al*, Zhao *et al*, and Liu and Koblan seeks to facilitate the generation of the AP site (middle). In other words, this study employs proteins that repair and not maintain/generate abasic sites whereas other studies employ proteins that generate/maintain and not repair abasic sites. CGBEs were designed based on working hypotheses derived from known Cas9 and BER chemistries.

[0031] **Figure 13** shows 1 column graph that illustrates C:G editing in A549 cells using various gRNA:target site pairs. (n=2, error bars denote standard error). These cells were transfected via lipofection.

[0032] **Figure 14** shows 1 column graph that illustrates C:G editing in HTB-9 cells using various gRNA:target site pairs. (n=2 error bars denote standard error). CGBEs are able to efficiently induce C:G to G:C edits at HEK2-1, HEK4, RNF2-1, and VEGFA. Additionally, C6 editing (RNF2-3) appears to be higher than C5 editing (RNF2-2). Collectively, the data indicate that the editing preferences of CGBEs in HEK cells carry over to HTB9 cells. However, the more efficient CGBE in HTB9 cells is ACX, rPB(8kD). *p < 0.05; **p < 0.01 (two tailed Student's T-test against 'Untreated,' n=2, mean ± standard error). Figure 14 illustrates ACX, rXRCC1; ACX, rPB(8kD); and BE3 editing in HTB9 cells.

[0033] **Figure 15** shows 1 column graph that illustrates C:G editing in CRL5868 cells using various gRNA:target site pairs (n=1). These cells were transfected via nucleofection.

[0034] **Figure 16** shows 1 column graph that illustrates C:G editing in CAMA-1 cells using various gRNA:target site pairs (n=1). These cells were transfected via nucleofection.

[0035] **Figure 17** shows 1 column graph that illustrates C:G editing eHAP cells using various gRNA:target site pairs (n=3). Although BE3 editing is low, we observed moderate levels of editing with both CGBEs. These results suggest that CGBE may be able to induce some C:G to G:C edits in certain circumstances under which similar base editing technology – like BE3 – may not be as efficient (C:G to T:A edits; light blue). *p < 0.05; **p < 0.01; ***p < 0.001 (two tailed Student's T-test against 'Untreated,' n=3, mean ± standard error). Figure 17 illustrates ACX, rXRCC1; ACX, rPB(8kD); and BE3 editing in eHAP cells.

[0036] Figure 18 shows 1 column graph that illustrates C:G editing in HepG2 cells using various gRNA:target site pairs (n=1). These cells were transfected via nucleofection.

[0037] Figure 19 shows 1 column graph that illustrates C:G editing in Jurkat cells using various gRNA:target site pairs (n=1). These cells were transfected via nucleofection.

[0038] Figure 20 shows 1 column graph that illustrates C:G editing in H9 stem cells using various gRNA:target site pairs (n=3, error bars denote standard error). Without further engineering (via codon optimization, APOBEC mutation etc.), BE3 is inefficient at inducing C:G to T:A edits in H9 stem cells. The highest C:G to T:A editing observed with BE3 is at HEK4 (1.2%). Similarly, both CGBEs are not efficient at inducing C:G to G:C edits, with the highest edits also at HEK4. It was recently shown that the engineered human APOBEC3A22 can increase BE3 editing in stem cells²³, suggesting that a similar approach might also induce higher CGBE stem cell editing. *p < 0.05; **p < 0.01; ***p < 0.001 (two tailed Student's T-test against 'Untreated,' n=3, mean ± standard error). Figure 20 illustrates ACX, rXRCC1; ACX, rPB(8kD); and BE3 exhibit low editing efficiencies in H9 stem cells.

GENERAL DEFINITIONS

[0039] Several terms that are employed throughout the specification are generally defined in the following paragraphs. Other definitions may also found within the body of the specification.

[0040] As used herein, the terms "about" and "approximately," in reference to a number, is used herein to include numbers that fall within a range of 20%, 10%, 5%, 2.5%, 2%, 1.5% or 1% in either direction (greater than or less than) of the number unless otherwise stated or otherwise evident from the context (except where such number would exceed 100% of a possible value).

[0041] The terms "polynucleotide", "nucleic acid" and "oligonucleotide" are used interchangeably and refer to a polymeric form of nucleotides of any length, either deoxyribonucleotides or ribonucleotides or analogues thereof. Polynucleotides can have any three-dimensional structure and may perform any function, known or unknown. The following are non-limiting examples of polynucleotides: a gene or gene fragment (for example, a probe, primer, EST or SAGE tag), exons, introns, messenger RNA (mRNA), transfer RNA, ribosomal RNA, ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes and primers. A polynucleotide can comprise modified nucleotides, such as methylated nucleotides and nucleotide analogues. If present, modifications to the nucleotide structure can be imparted before or after assembly of the polynucleotide. The sequence of nucleotides can be interrupted by non-nucleotide components. A polynucleotide can be further modified after

polymerization, such as by conjugation with a labelling component. The term also refers to both double- and single-stranded molecules. Unless otherwise specified or required, a polynucleotide encompasses both the double-stranded form and each of two complementary single-stranded forms known or predicted to make up the double-stranded form. As used herein, the term "polypeptide" generally has its art-recognized meaning of a polymer of amino acids. The term is also used to refer to specific functional classes of polypeptides, such as, for example, nucleases, antibodies, etc.

[0042] As used herein, the term "variant" refers to an entity that shows significant structural identity with a reference entity but differs structurally from the reference entity in the presence or level of one or more chemical moieties as compared with the reference entity. In many embodiments, a variant also differs functionally from its reference entity. In general, whether a particular entity is properly considered to be a "variant" of a reference entity is based on its degree of structural identity with the reference entity. As will be appreciated by those skilled in the art, any biological or chemical reference entity has certain characteristic structural elements. A variant, by definition, is a distinct chemical entity that shares one or more such characteristic structural elements. To give but a few examples, a polypeptide may have a characteristic sequence element comprising a plurality of amino acids having designated positions relative to one another in linear or three-dimensional space and/or contributing to a particular biological function; a nucleic acid may have a characteristic sequence element comprised of a plurality of nucleotide residues having designated positions relative to one another in linear or three-dimensional space. For example, a variant polypeptide may differ from a reference polypeptide as a result of one or more differences in amino acid sequence and/or one or more differences in chemical moieties (e.g., carbohydrates, lipids, etc.) covalently attached to the polypeptide backbone. In some embodiments, a variant polypeptide shows an overall sequence identity with a reference polypeptide (e.g., a nucleic acid modifying enzyme described herein) that is at least 60%, 65%, 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, or 99%. Alternatively or additionally, in some embodiments, a variant polypeptide does not share at least one characteristic sequence element with a reference polypeptide. In some embodiments, the reference polypeptide has one or more biological activities. In some embodiments, a variant polypeptide shares one or more of the biological activities of the reference polypeptide, e.g., enzymatic activity. In some embodiments, a variant polypeptide lacks one or more of the biological activities of the reference polypeptide. In some embodiments, a variant polypeptide shows a reduced level of one or more biological activities (e.g., enzymatic activity) as compared with the reference polypeptide. In some embodiments, a polypeptide of interest is considered to be a "variant" of a parent or reference polypeptide if the polypeptide of interest has an amino acid sequence

that is identical to that of the parent but for a small number of sequence alterations at particular positions. Typically, fewer than 20%, 15%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2% of the residues in the variant are substituted as compared with the parent. In some embodiments, a variant has 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1 substituted residue as compared with a parent. Often, a variant has a very small number (e.g., fewer than 5, 4, 3, 2, or 1) number of substituted functional residues (i.e., residues that participate in a particular biological activity). Furthermore, a variant typically has not more than 5, 4, 3, 2, or 1 additions or deletions, and often has no additions or deletions, as compared with the parent. Moreover, any additions or deletions are typically fewer than about 25, about 20, about 19, about 18, about 17, about 16, about 15, about 14, about 13, about 10, about 9, about 8, about 7, about 6, and commonly are fewer than about 5, about 4, about 3, or about 2 residues. In some embodiments, the parent or reference polypeptide is one found in nature.

[0043] As used herein, the term "expression" of a nucleic acid sequence refers to the generation of any gene product from the nucleic acid sequence. In some examples, a gene product can be an RNA transcript. In some embodiments, a gene product can be a polypeptide. In some embodiments, expression of a nucleic acid sequence involves one or more of the following: (1) production of an RNA template from a DNA sequence (e.g., by transcription); (2) processing of an RNA transcript (e.g., by splicing, editing, 5' cap formation, and/or 3' end formation); (3) translation of an RNA into a polypeptide or protein; and/or (4) post-translational modification of a polypeptide or protein.

[0044] Unless specified otherwise, all amino acid sequences are shown with the NH₂ end on the left and the COOH end on the right, and all DNA/RNA acid sequences are shown with the 5' end on the left and the 3' end on the right.

DETAILED DESCRIPTION

[0045] Disclosed herein are compositions and methods of editing a nucleobase, for example, generating a cytosine to guanine mutation (or conversion) in a polynucleotide. The inventors have developed a new class of C:G to G:C Base Editors (CGBEs) which utilize or manipulate the Base Excision Repair (BER) pathway downstream of abasic site creation. When a cytosine to guanine conversion is effected on one strand of a double stranded DNA polynucleotide, the opposing guanine on the opposing strand may also be converted into a cytosine by the intrinsic DNA repair mechanisms. Therefore, this new class of CGBEs edits C:G to G:C (Figure. 1a), which opens up treatment avenues to 11% (singular CGBE) to 40% (CGBE with cytosine base editors/adenine base editors (CBE/ABE) disease-associated single-nucleotide polymorphisms (SNPs) (Table 1).

[0046] Table 1. CGBEs enable potential treatment avenues to previously addressable SNPs associated with human diseases. CBE enables treatment to 48% of all known disease-associated SNPs, while adenine base editors (ABE) enables treatment to 6%. CGBEs effect primarily C:G to G:C and G:C to C:G changes (*Row 6 in Table 1*) that can correct 11% of disease-associated SNPs. In combination with cytosine base editors (CBEs) or ABEs, CGBEs effect secondarily G to T, C to A, A to C, and T to G edits (*Rows 3 and 5 in Table 1*). With CBEs, ABEs, and CGBEs, the remaining 7% of SNPs (A to T and T to A) can also be corrected (*Row 1 in Table 1*).

Row	From	To	% of SNPs in ClinVar	Equivalent to		Base-editing route for correction
1	A	T	7	T	A	A to G (ABE), G to C (CGBE), C to T (CBE)
2		G	48		C	ABE
3		C	15		G	C to G (CGBE), G to A (CBE)
4	G	A	6	C	T	CBE
5		T	14		A	A to G (ABE), G to C (CGBE)
6		C	11		G	CGBE

[0047] In one aspect, the present disclosure refers to a fusion protein or a protein complex comprising a DNA binding protein (DnaBP), a nucleobase modifying protein (NMP), and a Base Excision Repair associated protein (BERAP); wherein the fusion protein or protein complex does not comprise a Uracil binding protein or a catalytically active DNA polymerase.

[0048] In this aspect, the nucleobase editor comprises at least three components: a DNA binding protein (DnaBP), a nucleobase modifying protein (NMP), and a Base Excision Repair associated protein (BERAP). The nucleobase editor can be a fusion protein (a single polypeptide translated from a fusion gene) or a protein complex. As used herein, the term "protein complex" refers to a composite unit that is a combination of two or more proteins formed by interaction between the proteins. Typically but not necessarily, a "protein complex" is formed by the binding of two or more proteins together through specific non-covalent binding affinities. However, covalent bonds may also be present between the interacting partners. For instance, the two interacting partners can be covalently crosslinked so that the protein complex becomes more stable.

[0049] *DNA binding proteins (DnaBP)*

[0050] The term "DNA binding protein (DnaBP)" refers to a protein which is capable of binding with a DNA. In some examples, the DNA binding protein is a programmable DNA binding protein, which can be designed or programmed to bind with a specific DNA sequence. In some examples, the programmable DNA binding protein is an RNA-guided DNA binding protein. As used herein, an RNA-guided DNA binding protein interacts or forms a complex with a guide RNA, and can specifically target or bind with a polynucleotide of a specific

sequence which usually comprises a sequence complementary to the targeting domain of the gRNA. Upon binding with the target polynucleotide, the DNA binding protein may remain bound with the target polynucleotide, or it may modify the target polynucleotide. In one example, the DNA binding protein is a CRISPR-associated protein (Cas). Many Cas proteins possess endonuclease activity and are also termed Cas nucleases. In a specific example, the DNA binding protein is a Cas protein. In some examples, the Cas protein is selected from the group including but not limited to Cas3, a Cas9, a xCas9, a SpRY Cas9, a HF-Cas9, a Cas9-NG, a circularly permuted Cas9, a codon-optimised Cas9, a domain-fused Cas9, a Cas10 and a Cas12 (also known as Cpf1), a Cas14, a CasX, a CasΦ, and variants thereof. In some examples, the DnaBP is a nickase variant of any of the Cas proteins aforementioned. Therefore in some examples, the Cas domain is a Cas nickase (nCas). In some examples, the DnaBP is a Cas9 nickase (or nCas9). The term "Cas9 nickase," as used herein, refers to a Cas9 protein that is capable of cleaving only one strand of a duplexed nucleic acid molecule (e.g., a duplexed DNA molecule). In some examples, the Cas domain is a nuclease inactive Cas (dCas).

[0051] The terms "guide RNA" and "gRNA" refer to any nucleic acid that promotes the specific association (or "targeting") of a DNA binding protein to a target sequence either in a cell or in a cell free environment. gRNAs can be unimolecular (comprising a single RNA molecule, and referred to alternatively as chimeric), or modular (comprising more than one, and typically two, separate RNA molecules, such as a crRNA and a tracrRNA, which are usually associated with one another, for instance by duplexing).

[0052] *Base Excision Repair associated proteins (BERAP)*

[0053] As used herein, the term "Base Excision Repair associated protein (BERAP)" refers to any protein that is involved in the Base Excision Repair pathway. BERAP may also be referred to as "BER protein". In some examples, the BERAP is an enzyme functioning in one or more steps of the BER pathway; in other examples, the BERAP is a co-factor or a scaffold protein of enzymes functioning in the BER pathway. Scaffold proteins are understood to be proteins which regulate the function or activity of other proteins or pathways by interacting or binding with one or more members of the pathways. In some cases, a scaffold protein may tether multiple members of a pathway into complexes. BERAPs can be identified by referring to proteins listed in the KEGG (Kyoto Encyclopedia of Genes and Genomes) BER pathway (https://www.genome.jp/kegg-bin/show_pathway?map=ko03410), or by searching the term 'base excision repair' in protein databases such as Uniprot (www.uniprot.org).

[0054] In some examples of the fusion protein or protein complex as disclosed herein, the BERAP is selected from the group including but not limited to: an AP endonuclease, an end

processing enzyme, a catalytically inactive DNA polymerase, a lyase domain, a Flap endonuclease, a DNA ligase, and a scaffold protein involved in the BER pathway. In some examples, the BERAP is selected from the group including but not limited to: a DNA ligase III (LIG3), an XRCC1, a DNA binding or lyase domain of DNA Polymerase beta (PB), a DNA binding or lyase domain of DNA Polymerase delta, a DNA binding or lyase domain of DNA Polymerase epsilon, an AP endonuclease (APE1), Proliferating cell nuclear antigen (PCNA), DNA-(apurinic or apyrimidinic site) lyase (APEX), Poly (ADP-ribose) polymerase (PARP), Flap endonuclease 1 (FEN1), and DNA ligase I (LIG1). In one example, the BERAP is an XRCC1. In some examples, the BERAP is a rat XRCC1 (rXRCC1) or a variant thereof. In some examples, the BERAP is a human XRCC1 (hXRCC1) or a variant thereof. In one example, the BERAP is a rXRCC1 with the amino acid sequence of SEQ ID NO: 4. In one example, the BERAP is a hXRCC1 with the amino acid sequence of SEQ ID NO: 5.

[0055] In some examples of the fusion protein or protein complex as disclosed herein, the BERAP is a DNA binding or lyase domain of DNA Polymerase beta (POLB, or PB). In some specific examples, the DNA binding or lyase domain of DNA Polymerase beta corresponds to a region contained within amino acids 1-140, 1-120, 1-100, or 1-87 of the full DNA Polymerase beta sequence. In one example, the DNA binding or lyase domain of DNA Polymerase beta corresponds to a region contained within amino acids 1-140, 1-120, 1-100, or 1-87 of the full human DNA Polymerase beta sequence (SEQ ID NO: 12). In another example, the DNA binding or lyase domain of DNA Polymerase beta corresponds to a region contained within amino acids 1-140, 1-120, 1-100, or 1-87 of the full rat DNA Polymerase beta sequence (SEQ ID NO: 13). In some examples, the BERAP is a human DNA binding or lyase domain of DNA Polymerase beta (PB) or a variant thereof. In some examples, the BERAP is a rat DNA binding or lyase domain of DNA Polymerase beta (PB) or a variant thereof. In one example, the BERAP is a DNA binding or lyase domain of rat Polymerase beta (rPB), with an amino acid sequence of SEQ ID NO: 6. In one example, the BERAP is a DNA binding or lyase domain of rat Polymerase beta (rPB), with an amino acid sequence of SEQ ID NO: 7. In another example, the BERAP is a DNA binding or lyase domain of human Polymerase beta (hPB), with an amino acid sequence of SEQ ID NO: 8. In another example, the BERAP is a DNA binding or lyase domain of human Polymerase beta (hPB), with an amino acid sequence of SEQ ID NO: 9.

[0056] In some examples of the fusion protein or protein complex as disclosed herein, the BERAP is a DNA Ligase III (LIG3). In one example, the BERAP is a rat DNA Ligase III (LIG3), with an amino acid sequence of SEQ ID NO: 10. In another example, the BERAP is a human DNA Ligase III (LIG3), with an amino acid sequence of SEQ ID NO: 11.

[0057] *Nucleobase modifying proteins (NMP)*

[0058] The term “nucleobase modifying protein (NMP)” refers to any protein domain that is capable of modifying a nucleobase (such as adenine (A), cytosine (C), guanine (G), thymine (T), and uracil (U)). The modification can be any chemical or physical changes to the nucleobase, and the NMP includes but not limited to a deaminase, a nuclease, a nickase, a recombinase, a methyltransferase, a methylase, an acetylase, and an acetyltransferase.

[0059] In some examples of the fusion protein or protein complex as disclosed herein, the nucleobase modifying protein (NMP) is a cytosine deaminase domain. A cytosine deaminase domain is the functional domain of a cytosine deaminase that has deaminase activity. As the deamination occurs on the cytosine nucleobase that is comprised in a cytidine or deoxycytidine, a cytidine deaminase may also be referred to as a cytosine deaminase. In the present disclosure, the terms “cytidine deaminase” is used interchangeably with “cytosine deaminase”. For example, the apolipoprotein B mRNA-editing complex (APOBEC) family of deaminases are conventionally referred to as cytidine deaminases, but they are capable of deaminating the cytosine in cytidine or deoxycytidine. Accordingly, the deaminase domain from an APOBEC cytidine deaminase protein can also be referred to as a cytosine deaminase domain. Therefore in some examples, the cytidine deaminase domain is a deaminase domain from an apolipoprotein B mRNA-editing complex (APOBEC) family deaminase. In some examples, the APOBEC family deaminase is selected from the group consisting of APOBEC1 deaminase, APOBEC2 deaminase, APOBEC3A deaminase, APOBEC3B deaminase, APOBEC3C deaminase, APOBEC3D deaminase, APOBEC3F deaminase, APOBEC3G deaminase, APOBEC3H deaminase, and any derivatives thereof. In some examples, the cytidine deaminase domain is an activation-induced deaminase (AID). In some examples, the cytidine deaminase domain is a cytidine deaminase 1 from *Petromyzon marinus* (pmCDA1). In some examples, the cytidine deaminase domain has a higher activity on methylated Cs. In some examples, the cytidine deaminase domain has a narrower targeting window.

[0060] *The exclusion of uracil binding protein*

[0061] The fusion protein or a protein complex as disclosed herein does not comprise a Uracil binding protein. As used herein, the term “uracil binding protein” or “UBP” refers to a protein that is capable of binding to uracil. In some examples, the UBP is a uracil modifying enzyme, a uracil base excision enzyme or a uracil DNA glycosylase (UDG or UNG). Therefore, while uracil DNA glycosylase is considered to be involved in the BER pathway and is responsible for removing the creating the abasic site, it is not comprised in the fusion protein or a protein complex (the CGBE) as defined in claim 1. Without being bound by theory, the presence of a uracil binding protein (such as UDG) as a component of the CGBE may maintain the abasic site (created by the removal of the uracil base) and hinder the downstream BER pathway which repairs the abasic site. The CGBE of the present disclosure is designed to

promote the Base Excision Repair (BER) pathway, it does not comprise a uracil binding protein.

[0062] *Architecture of the CGBEs*

[0063] In some examples of the fusion protein or a protein complex as disclosed herein, the DNA binding protein (DnaBP) is a nickase Cas protein such as nCas9; the Base Excision Repair associated protein (BERAP) is selected from the group consisting of an XRCC1, a DNA Ligase III (LIG3), and a DNA binding or lyase domain of DNA Polymerase beta; and the nucleobase modifying protein (NMP) is a deaminase domain from an apolipoprotein B mRNA-editing complex (APOBEC) family deaminase. In some examples, the DNA binding protein (DnaBP) is a nickase Cas9 (nCas9); the Base Excision Repair associated protein (BERAP) is an XRCC1; and the nucleobase modifying protein (NMP) is a deaminase domain from an apolipoprotein B mRNA-editing complex (APOBEC) family deaminase. In some examples, the DNA binding protein (DnaBP) is a nickase Cas9 (nCas9); the Base Excision Repair associated protein (BERAP) is a DNA binding or lyase domain of DNA Polymerase beta; wherein the nucleobase modifying protein (NMP) is a deaminase domain from an apolipoprotein B mRNA-editing complex (APOBEC) family deaminase. In some examples, the DNA binding protein (DnaBP) is a nickase Cas9 (nCas9); the Base Excision Repair associated protein (BERAP) is a DNA Ligase III (LIG3); and the nucleobase modifying protein (NMP) is a deaminase domain from an apolipoprotein B mRNA-editing complex (APOBEC) family deaminase.

[0064] Different architectures or orientations of the fusion protein as disclosed herein are possible. In some examples of the fusion protein as disclosed herein, the orientation of the DnaBP, NMP and BERAP within the fusion protein is selected from the group consisting of: [NMP]-[DnaBP]-[BERAP], [NMP]-[BERAP]-[DnaBP] and [BERAP]-[NMP]-[DnaBP]; wherein each instance of "[]-" comprises an optional linker. In one example, the orientation of the DnaBP, NMP and BERAP within the fusion protein is [NMP]-[DnaBP]-[BERAP]. In one example, the orientation of the DnaBP, NMP and BERAP within the fusion protein is [NMP]-[DnaBP]-[BERAP].

[0065] The term "linker," as used herein, refers to a bond (e.g., covalent bond), chemical group, or a molecule linking two molecules or moieties, e.g., two domains of a fusion protein, such as, for example, a nuclease-inactive Cas9 domain and a nucleic acid-editing domain (e.g., an adenosine deaminase). In some embodiments, a linker joins a gRNA binding domain of an RNA-programmable nuclease, including a Cas9 nuclease domain, and the catalytic domain of a nucleic-acid editing protein. In some embodiments, a linker joins a dCas9 and a nucleic-acid editing protein. Typically, the linker is positioned between, or flanked by, two groups, molecules, or other moieties and connected to each one via a covalent bond, thus

connecting the two. In some embodiments, the linker is an amino acid or a plurality of amino acids (e.g. a peptide or protein). In some embodiments, the linker is an organic molecule, group, polymer, or chemical moiety. In some embodiments, the linker is 5-100 amino acids in length, for example, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-150, or 150-200 amino acids in length. Longer or shorter linkers are also contemplated. In some embodiments, a linker comprises the amino acid sequence SGSETPGTSESATPES (SEQ ID NO: 101), which may also be referred to as the XTEN linker. In some embodiments, a linker comprises the amino acid sequence SGGG (SEQ ID NO: 102). In some embodiments, a linker comprises SGGSGGGG (SEQ ID NO: 103), GGGG (SEQ ID NO: 104), G, EAAAK (SEQ ID NO: 105), GGS, SGSETPGTSESATPES (SEQ ID NO: 101) or XP motif, or a combination of any of these. In some embodiments, a linker comprises repeats of SGGSGGGG (SEQ ID NO: 103), GGGG (SEQ ID NO: 104), G, EAAAK (SEQ ID NO: 105), GGS, or XP, wherein the repeats is denoted by n, wherein n is independently an integer between 1 and 30, and wherein X is any amino acid. In some embodiments, n is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15.

[0066] In another aspect, the present disclosure provides a fusion protein comprising:

- a first amino acid sequence that is at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, or 99% identical to the amino acid sequence of any one of SEQ ID NOs: 1-2,
- a second amino acid sequence that is at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, or 99% identical to the amino acid sequence of SEQ ID NO: 3,
- a third amino acid sequence that is at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, or 99% identical to the amino acid sequence of SEQ ID NO: 4-11.

[0067] In one example of the fusion protein as disclosed herein, the fusion protein comprises one of the below structures:

[second amino acid sequence]-[first amino acid sequence]-[third amino acid sequence];
 [second amino acid sequence]-[third amino acid sequence]-[first amino acid sequence];
 [first amino acid sequence]-[second amino acid sequence]-[third amino acid sequence];
 [first amino acid sequence]-[third amino acid sequence]-[second amino acid sequence];
 [third amino acid sequence]-[first amino acid sequence]-[second amino acid sequence];
 [third amino acid sequence]-[second amino acid sequence]-[first amino acid sequence];
 wherein each instance of "-" comprises an optional linker.

[0068] In one example of the fusion protein as disclosed herein, any linker is independently 1-50 amino acids in length. In one example of the fusion protein as disclosed herein, any linker

is independently 1-25 amino acids in length. In one example of the fusion protein as disclosed herein, any linker is independently 5-20 amino acids in length. In one example of the fusion protein as disclosed herein, any linker is independently 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 amino acids in length.

[0069] In one example of the fusion protein as disclosed herein, any linker independently comprises one or more amino acid sequences selected from the group consisting of: SGSETPGTSESATPES (XTEN linker) (SEQ ID NO:101, SGGGS (SEQ ID NO:102) and GGGGS (SEQ ID NO:104).

[0070] In another aspect, the present disclosure provides a protein complex comprising:

- a first protein comprising an amino acid sequence that is at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, or 99% identical to the amino acid sequence of any one of SEQ ID NOs: 1-2,
- a second protein comprising an amino acid sequence that is at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, or 99% identical to the amino acid sequence of SEQ ID NO: 3,
- a third protein comprising an amino acid sequence that is at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, or 99% identical to the amino acid sequence of SEQ ID NO: 4-10.

[0071] In some examples, the fusion protein as disclosed herein comprises an amino acid sequence selected from the group consisting of: SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, and SEQ ID NO:72. In one example, the fusion protein as disclosed herein has an amino acid sequence of SEQ ID NO: 44. In another example, the fusion protein as disclosed herein has an amino acid sequence of SEQ ID NO: 48. In another example, the fusion protein as disclosed herein has an amino acid sequence of SEQ ID NO: 69. In another example, the fusion protein as disclosed herein has an amino acid sequence of SEQ ID NO: 70.

[0072] In another aspect, the present disclosure provides a protein-nucleic acid complex comprising a nucleic acid molecule and any one of:

- the fusion protein or protein complex as disclosed herein,
- the fusion protein as disclosed herein; and
- the protein complex as disclosed herein.

[0073] In one example of the protein-nucleic acid complex as disclosed herein, the nucleic acid molecule is an RNA. In some examples, the RNA is a guide RNA (gRNA), or more specifically a single guide RNA (sgRNA). In some examples, the single guide RNA (sgRNA) as disclosed herein comprises a sequence selected from the group consisting of: SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40 and SEQ ID NO:41.

[0074] As used herein, the term “protein-nucleic acid complex” means a complex unit that is a combination of at least one protein and at least one nucleic acid formed by an interaction, including an interaction between the protein and the nucleic acid. Typically, “protein-nucleic acid complexes” are formed by, but not necessarily, the binding of proteins and nucleic acids through non-covalent affinity. In some examples, the gene editing complex is a protein-nucleic acid complex such as a ribonucleoprotein (RNP). A non-limiting example of an RNP is CRISPR-Cas RNP that includes a Cas protein and gRNA.

[0075] In one example of the protein-nucleic acid complex as disclosed herein, the nucleic acid molecule comprises a sequence which is about 80%, 90%, or 95% identical or reverse complementary to any of the target sequences listed in Table 2. In some examples, the nucleic acid molecule comprises a sequence which is identical or reverse complementary to any of the target sequences listed in Table 2.

[0076] **Table 2.** Target protospacer sequences of exemplary guide RNAs which are used to effect C to G conversions in disease associated genes. Targeted C's are underlined. PAMs are in bold.

SEQ ID NO.	Name	Target Sequences and PAMs
73	EMX1	GAGT <u>CC</u> GAGCAGAAGAAGAAGGG
74	FANCF	GGAAT <u>CC</u> CTTCTGCAGCACCTGG
75	HEK2-1	GAACA <u>C</u> AAAGCATAGACTGCGGG
76	HEK2-2	GAGCA <u>CC</u> CACACCCTAAACTATGG
77	HEK2-3	GGAA <u>A</u> CGGATAGTTCTGAAAGGG
78	HEK2-4	CTTAA <u>C</u> TATTTGTATTCCACTGG
79	HEK2-5	CTTCC <u>A</u> AGTGAGAAGCCAGTGG
80	HEK2-6	CCAGC <u>CC</u> GCTGGCCCTGTAAAGG
81	HEK2-7	CATT <u>C</u> GTTATTTTACATATTGG
82	HEK2-8	GTTTC <u>C</u> TTTACAGGGCCAGCGGG

83	HEK2-9	ATACG <u>C</u> ACAGTTTGACAGAT G G G
84	HEK2-10	GCTGG <u>C</u> CCTGTAAAGGAA A C T G G
85	HEK2-11	GCA <u>T</u> G <u>C</u> GTGTGTGTTTAAG C T G G
86	HEK2-12	TTGGG <u>C</u> TGCAGTAACTTGA A G G G
87	HEK2-13	TCTTT <u>C</u> AAGCAGGTGATTAC A G G
88	HEK2-14	AGTTT <u>C</u> CTTACAGGGCCAG C G G
89	HEK2-15	GAGGT <u>C</u> GTGGCTGAGCACA A G G G
90	HEK2-16	GGCCT <u>C</u> TATTGTTGGTAGAAT G G
91	HEK3	GGC <u>C</u> CAGACTGAGCACGTGAT G G
92	HEK4	GGCA <u>C</u> TGCGGCTGGAGGTGG G G G
93	RNF2-1	GTCAT <u>C</u> TTAGTCATTACCTG A G G
94	RNF2-2	<u>C</u> AC <u>A</u> <u>C</u> AC <u>A</u> CTTAGAATCTGT G G G
95	RNF2-3	A <u>C</u> A <u>C</u> A <u>C</u> AC <u>A</u> CTTAGAATCTGT G G
96	GJB2	GGACA <u>C</u> GAAGATCAGCTGC A G G G
97	ADRB2	CCCTTT <u>C</u> CTGCGTGACGTCG T G G
98	MYBPC3	CCCTTT <u>C</u> CTGCGTGACGTCG T G G
99	GAL 292	GAAGT <u>C</u> GTTGTCAAACAGGA A G G
100	VEGFA-1	GATGT <u>C</u> TGCAGGCCAGATG A G G G

[0077] In another aspect, the present disclosure provides a method of replacing a cytosine with a guanine on a DNA strand in a cell, said method comprises introducing to the cell the fusion protein or protein complex as disclosed herein, the fusion as disclosed herein, the protein complex as disclosed herein, or the protein-nucleic acid complex as disclosed herein.

[0078] In some examples, the cell is a eukaryotic cell. In some examples, the cell is an animal cell. In some specific examples, the cell is a human cell. In some examples, the method is performed *in vivo* or *in vitro*.

[0079] In some examples, the BERAP and the NMP interact with the same strand of a target DNA molecule.

[0080] In another aspect, the present disclosure provides a vector comprising the polynucleotide as disclosed herein.

[0081] In another aspect, the present disclosure provides a kit comprising the fusion protein or protein complex as disclosed herein, the fusion protein as disclosed herein, the protein complex as disclosed herein, or the protein-nucleic acid complex as disclosed herein.

[0082] In another aspect, the present disclosure provides a cell comprising the fusion protein or protein complex as disclosed herein, the fusion protein as disclosed herein, the protein complex as disclosed herein, or the protein-nucleic acid complex as disclosed herein.

[0083] In another aspect, the present disclosure provides a cell comprising one or more nucleic acid molecules that encode the fusion protein or protein complex as disclosed herein, the fusion protein as disclosed herein, the protein complex as disclosed herein, or the protein-nucleic acid complex as disclosed herein.

[0084] In another aspect, the present disclosure provides a method of treating a subject having or suspected of having a disease or disorder comprising administering the fusion protein or protein complex as disclosed herein, the fusion protein as disclosed herein, the protein complex as disclosed herein, or the protein-nucleic acid complex as disclosed herein, the pharmaceutical composition as disclosed herein, the polynucleotide as disclosed herein, or the vector as disclosed herein to the subject. In one example, the disease or disorder comprises one or more C to G (C>G) mutations. In another example, the disease or disorder comprises one or more G to C (G>C) mutations. In another example, the disease or disorder comprises C>G and G>C mutations. In one example, the disease or disorder is inclusive of, but not limited to any of the conditions listed in <https://www.ncbi.nlm.nih.gov/clinvar/?term=C%3EG> or <https://www.ncbi.nlm.nih.gov/clinvar/?term=G%3EC>, and filtering based on the selection of the following clinical significance: pathogenic, risk factor or likely pathogenic. In one example, the disease or disorder is selected from a group consisting of skin fibrosis, bladder cancer, liver cancer, Myasthenic syndrome, Spondyloepimetaphyseal dysplasia, Parkinson's disease, Deafness, blood disorders, and Schnyder crystalline corneal dystrophy.

[0085] In another aspect, the present disclosure provides a fusion protein or protein complex as disclosed herein, the fusion protein as disclosed herein, the protein complex as disclosed herein, or the protein-nucleic acid complex as disclosed herein, the pharmaceutical composition as disclosed herein, the polynucleotide as disclosed herein, or the vector as disclosed herein for use in treating a subject having or suspected of having a disease or disorder. In one example, the disease or disorder comprises one or more C to G (C>G) mutations. In another example, the disease or disorder comprises one or more G to C (G>C) mutations. In another example, the disease or disorder comprises C>G and G>C mutations. In one example, the disease or disorder is inclusive of, but not limited to any of the conditions listed in <https://www.ncbi.nlm.nih.gov/clinvar/?term=C%3EG> or <https://www.ncbi.nlm.nih.gov/clinvar/?term=G%3EC>, and filtering based on the selection of the following clinical significance: pathogenic, risk factor or likely pathogenic. In one example, the disease or disorder is selected from a group consisting of skin fibrosis, bladder cancer,

liver cancer, Myasthenic syndrome, Spondyloepimetaphyseal dysplasia, Parkinson's disease, Deafness, blood disorders, and Schnyder crystalline corneal dystrophy.

[0086] In another aspect, the present disclosure provides a fusion protein or protein complex as disclosed herein, the fusion protein as disclosed herein, the protein complex as disclosed herein, or the protein-nucleic acid complex as disclosed herein, the pharmaceutical composition as disclosed herein, the polynucleotide as disclosed herein, or the vector as disclosed herein in the manufacture of a medicament for treating a subject having or suspected of having a disease or disorder. In one example, the disease or disorder comprises one or more C to G (C>G) mutations. In another example, the disease or disorder comprises one or more G to C (G>C) mutations. In another example, the disease or disorder comprises C>G and G>C mutations. In one example, the disease or disorder is inclusive of, but not limited to any of the conditions listed in <https://www.ncbi.nlm.nih.gov/clinvar/?term=C%3EG> or <https://www.ncbi.nlm.nih.gov/clinvar/?term=G%3EC>, and filtering based on the selection of the following clinical significance: pathogenic, risk factor or likely pathogenic. In one example, the disease or disorder is selected from a group consisting of skin fibrosis, bladder cancer, liver cancer, Myasthenic syndrome, Spondyloepimetaphyseal dysplasia, Parkinson's disease, Deafness, blood disorders, and Schnyder crystalline corneal dystrophy.

[0087] In another aspect, the present disclosure provides a method for editing a target nucleobase pair of a double-stranded DNA sequence, the method comprising:

- a. contacting a target region of the double-stranded DNA sequence with a complex comprising a nucleobase editor and a guide nucleic acid, wherein the target region comprises the target nucleobase pair;
- b. inducing strand separation of said target region;
- c. converting a first nucleobase of said target nucleobase pair in a single strand of the target region to a second nucleobase;
- d. excising said second nucleobase from the double-stranded DNA sequence to produce an abasic site; and
- e. promoting the Base Excision Repair (BER) pathway to repair the abasic site, generating a third nucleobase at the abasic site, wherein the third nucleobase is different from the first nucleobase.

[0088] In one example, the method as disclosed herein further comprises converting a fourth nucleobase that is complementary to the third nucleobase, thereby generating an intended edited base pair.

[0089] In one example of the method as disclosed herein, the efficiency of generating the intended edited base pair is at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 35%, at least 40%, at least 45% or at least 50%.

[0090] In one example of the method as disclosed herein, the ratio of intended edited base pairs to unintended edited base pairs is at least 2:1, at least 3:1, at least 4:1, at least 5:1, at least 6:1, at least 7:1, at least 8:1, at least 9:1, or at least 10:1.

[0091] In one example of the method as disclosed herein, the first nucleobase is cytosine. In one example of the method as disclosed herein, the second nucleobase is uracil. In one example of the method as disclosed herein, the third nucleobase is guanine. In one example of the method as disclosed herein, the fourth nucleobase is cytosine.

[0092] In some examples of the method as disclosed herein, the nucleobase editor comprises nickase activity.

[0093] In some examples of the method as disclosed herein, the target region is 5-40, 5-30, 5-20, or 20 amino acids in length.

[0094] In some examples of the method as disclosed herein, the intended edited base pair resides within or proximal to the CGBE-binding site (protospacer). In some examples, the intended edited base pair is located 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides upstream, within, or downstream of the CBGE-binding site. In some examples, the intended edited base pair is upstream of a protospacer adjacent motif (PAM) site. In some examples, the intended edited base pair is located 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides upstream of the PAM site. As used herein, the protospacer adjacent motif (or PAM for short) is a short DNA sequence (usually 2-6 base pairs in length) that follows the DNA region targeted for cleavage by the CRISPR system, such as CRISPR-Cas9 (which has the PAM site sequence of NGG). In some examples, the intended edited base pair is located 14, 15, or 16 or 17 nucleotides upstream of the PAM site. Unless explicitly stated otherwise, the term "upstream of the PAM site" describes nucleotides/base pairs to the 5' direction of the PAM site, on the non-complementary strand (the strand not bound by the guide RNA).

[0095] In some examples of the method as disclosed herein, the intended edited base pair is downstream of a protospacer adjacent motif (PAM) site. In some examples, the intended edited base pair is located 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides downstream of the PAM site. As used herein, the protospacer adjacent motif (or PAM for short) is a short DNA sequence (usually 2-6 base pairs in length) that follows the DNA region targeted for cleavage by the CRISPR system, such as CRISPR-Cas9 (which has the PAM site sequence of NGG). In some examples, the intended edited base pair is located 14, 15, 16, 17, 18, 19 or 20 nucleotides downstream of the PAM site. Unless explicitly stated otherwise, the term "downstream of the PAM site" describes nucleotides/base pairs to the 3'

direction of the PAM site, on the non-complementary strand (the strand not bound by the guide RNA).

[0096] In some examples of the method as disclosed herein, the nucleobase editor comprises a linker. In some examples, the linker is 1-25, 5-20, 10-15 amino acids in length.

[0097] In some examples of the method as disclosed herein, the target region comprises a target window, wherein the target window comprises the target nucleobase pair. As used herein, the target region is a region on double-stranded DNA sequence which the fusion protein or protein complex (the CBGE) is designed to recognize or bind with. In examples wherein the DNA binding protein is guided by a guide RNA (such is the case for the Cas family proteins), the target region may be the region bound by the guide RNA. As used herein, the target window refers to a sequence window within the target region that is subject to efficient C to G editing of the CBGE. For optimal C to G editing, the target "C" is ideally located within the targeting window. In some examples, the target window is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides in length. In some examples, the target window is 3, 4, 5, 6, or 7 nucleotides in length. In some examples, the target window comprises the 14th, 15th, 16th and 17th nucleotides upstream of the PAM site. In one example, the target window is the 15th nucleotides upstream of the PAM site.

[0098] In some examples of the method as disclosed herein, the nucleobase editor comprises the fusion protein or protein complex as disclosed herein, the fusion protein as disclosed herein, or the protein complex as disclosed herein.

[0099] In some examples of the method as disclosed herein, the first nucleobase of the target nucleobase pair is a cytosine, and wherein said cytosine is in a DNA motif characterized by any one of the group consisting of WCW, ACC and GCT; wherein "C" is said Cytosine, W indicates an Adenine(A) or a Thymine (T).

[00100] In some examples, the method as disclosed herein is performed in vivo or in vitro.

[00101] Table 3. Sequence listing table

SEQ ID NO:	Name	Sequence
1	nCas9 (Cas9 nickase)	MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKN LIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE MAK VDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHL RKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLF IQLVQTYNQLFEENPINASGVDAKILSARLSKSRLENLIAQLPGE KKNGLFGNLIASLGLTPNFKSNFDLAEDAKLQLSKDTYDDLDNL LAQIGDQYADLFLAAKNLSDAILSDILRVNTEITKAPLSASMIKRYD EHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEE FYKFIKPILEKMDGTEELLVCLNREDLLRKQRTFDNGSIPHQIHLGE

		LHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWM TRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPK HSLLEYEFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTN RKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKIIKD KDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTIAHLFDDKVMKQ LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQM LIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVK VVDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEEG IKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL SDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK MKNYWRQLLNAKLITQRKFDNLTAKERGGLSELDKAGFIKRLVE TRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDF QFYKVINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVV DVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE TNGETGEIVWDKGRDFATVRKVLSPQVNVKKTEVQTGGFSKES ILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKS KKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYS LFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKG SPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVL SAY NKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKE VL DATLIHQSI TGLYETRIDLSQLGGD
2	Cas9 (Wildtype)	MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKN LIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE MAK VDDSFHRL EESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHL RKKLV DSTDKADLR LIY LALAHMIKFRGHFLIEGDLNPDNSDVKLF IQLVQTYNQLFEENPINASGVDAKAILSARLSKSRLENLIAQLPGE KKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDLDNL LAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYD EHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEE FYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGE LHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWM TRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPK HSLLEYEFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTN RKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKIIKD KDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTIAHLFDDKVMKQ LKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQM LIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVK VVDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRERMKRIEEG IKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL SDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKK MKNYWRQLLNAKLITQRKFDNLTAKERGGLSELDKAGFIKRLVE TRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDF QFYKVINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVV DVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE TNGETGEIVWDKGRDFATVRKVLSPQVNVKKTEVQTGGFSKES ILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKS KKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYS LFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKG SPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVL SAY NKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKE VL DATLIHQSI TGLYETRIDLSQLGGD
3	APOBEC	MSSETGPVAVDPTLRRRIEPHEFEVFFDPREL RKETCLLYEINWG GRHSIWRHTSQNTNKHVEVNFIEKFTTERTYFCPNTRCSITWFLSW SPCGECSRAITEFLSRYPHTLFIYIARLYHHADPRNRQGLRDLISS

		GVTIQIMTEQESGYCWRNRFVNYSNEAHWPYPHLLWVRLVYLE LYCIILGLPCLNLRKQPQLTFFTIALQSCHYQRLPPHILWATGLK
4	rXRCC1	MPEISLRHVSCSSQDSTHRAENLLKADTYRKWRSKAGEKTISV VLQLEKEEQIHSVDIGNDGSFAFVEVLVGSSAGGATAGEQDYEVLL VTSSFMSPSESRSRSGSNPNRVRIFGPDCLVRAAAEKRWDRVKIVCS QPYSKDSFYGLSFVKFHSPDPKDEAEAPSQKVTVTKLGGFRVKEE DDANSRLRPGALFFNRINKAASASASDPAGPSYAAATLQASSAAS SALPVPKVGSSSKLQEPKPKGRKLDLGLSDSKPPSKPSAGPAAL KRPKLPVPSRTPAATPASTPAQKAVPGKPRGEGTEPRGARAGPQ ELGKILQGVVVVLSGFQNPFRSELRDKALELGAKYRPDWTDPDSTH LICAFANTPKYSQVLGLGGRIVRKEWVLD CYRMRRRLPSRRYLMA GLGSSSEDEGDSHSESGEDEAPKLPRKRPQPKAKTQAAGPSSPP RPPTPEETKAPSPGPQDNDSDTDGEQSEGRDNGAEDSGDTEDEL RRVAKQREQRQPPAPEENGEDPYAGSTDENTDSEAPSEADLPI ELPDDFFQGHFFLYGEFPGDERRKLIRYVTA FNGELEDYMSDRVQ FVITAEWDPNFEEALMENPSLAFVRPRWIYSCNEKQKLLPHQLY GVVPQA
5	hXRCC1	MGPEIRLRHVSCSSQDSTHCAENLLKADTYRKWRAAKAGEKTIS VVLQLEKEEQIHSVDIGNDGSFAFVEVLVGSSAGGAGEQDYEVLLV TSSFMSPSESRSRSGSNPNRVRMFGPDCLVRAAAEKRWDRVKIVCS QPYSKDSFPGLSFVRFHSPDPKDEAEAPSQKVTVTKLGGFRVKEE DESANSRLRPGALFFSRINKTSPVTASDPAGPSYAAATLQASSAASS ASPVSRAGSTSKPQESPKGRKLDLNQEEKTPSKPPAQLSPSV PKRPKLPAPTRTPATAPV PARAQGA VTGKPRGEGTEPRRPRAGP EELGKILQGVVVVLSGFQNPFRSELRDKALELGAKYRPDWTDRST HLICAFANTPKYSQVLGLGGRIVRKEWVLDCHRMRRRLPSQRYLM AGPGSSSEDEASHSGSGDEAPKL PQKQPQT KTPTQAAGPSS PQKPPTPEETKAASPV LQEDIDIEGVQSEGQDNGAEDSGDTEDEL RRVAEQKEHRLPPGQEENGEDPYAGSTDENTDSEEHQEPDLPV PELPDDFFQGHFFLYGEFPGDERRKLIRYVTA FNGELEDYMSDRV QFVITAEWDPSFEEALMDNPSLAFVRPRWIYSCNEKQKLLPHQL YGVVPQA
6	rPB_8KD	MSKRKAPQETLNGGITDMLVELANFEKNVSQAIHKYNAYRKAASVI AKYPHKKISGAEAKKLPVGVTGKIAEKIDEFLATGKLRKLEK
7	rPB_14KD	MSKRKAPQETLNGGITDMLVELANFEKNVSQAIHKYNAYRKAASVI AKYPHKKISGAEAKKLPVGVTGKIAEKIDEFLATGKLRKLEKIRQDDT SSSINFLTRVTGIGPSAARKLVDEGIKTLEDLRKNEDKLNHHQRIGL K
8	hPBI_8KD	MSKRKAPQETLNGGITDMLTELANFEKNVSQAIHKYNAYRKAASVI AKYPHKKISGAEAKKLPVGVTGKIAEKIDEFLATGKLRKLEK
9	hPBI_14KD	MSKRKAPQETLNGGITDMLTELANFEKNVSQAIHKYNAYRKAASVI AKYPHKKISGAEAKKLPVGVTGKIAEKIDEFLATGKLRKLEKIRQDDT SSSINFLTRVSGIGPSAARKFVDEGIKTLEDLRKNEDKLNHHQRIGL K
10	rLIG3	MTLAFKTLFPRNLCALGRKELCLFSEQHHWPAIRQFSQWSETNLL CGCCLLQRRKPVLSFQRGHLRPRATHLISWSGSHVGLCTGPCEM AEQRFCDVYAKRGTAGCKKCKEKIVKGVCRIQKVVPNPFSESGG DMKEWYHIKCMFEKLERARATTKKIEDLTELEGWEELEDNEKEQI SQHIADLSSKTAATPKKKATVQAKLTTTGQVTSVPKVGASFITSTNP RKFSGFSAKPNNSEQDPSSPAPKTSLSASKCDPKHKDCLLREFR KLCAMVAENPSYNTKTQIIHDFLQKGSTGDGFRGDVYLTVKLLLP

		VIKSVYNLNDKQIVKLFSTRIFNCNPDDMARDLEQGDVSETIRVFFE QSKSFPPAAKSLLTIQEVDAFLHLSKLTKEDEQQQALQDIASRCT ANDLKCIIRLIKHDLKMNSGAKHVLDALDPNAYEAFKASRNLDQVV ERVLHNEQEVEKDPGRRRALSQASLMTVPQPMLEACKSIEYA MKKCPNGMFSEIKYDGERVQVHKKGDHFSYFSRSLKPVLPHKVA HFKDYIPKAFFGGQSMILDSEVLLIDNNTGKPLPFGTLGVHKKAAF QDANVCLFVFDCIYFNDVSLMDRPLCERRKFLHDNMVEIRNRMF SEMKQVTKASDLADMINRVIREGLEGLVLKDVKGTYEPGKRHWLK VKKDYLNEGAMADTADLVVLGAFYGGQSGKGMMSIFLMGCYDPD SQKWCTVTKCAGGHDDATLARLQKELDMVKISKDPSKIPSWLKIN KIYYPDFIVDPKKAHVWEITGAEFSRSEAHTADGISIRFPRCTRIR DDKDWKSATNLPQLKELYQLSKDKADFAVVAGDEGSSTTGGNSG ENEGTAGSTVPRKGPKGPPSKSSASAKKTEQKLNDPSSRGGEKL AVKSSPVKVGMRKKAADETPGLTKRRRRASRQRRRAMRTGRR
11	hLIG3	MSKAAGTPKKKAVVQAKLTTTGQVTSPVKGASFVTSTNPRKFSGF SAKPNNSGEAPSSPTPKRSLSSSKCDPRHKDCLLREFRKLKAMV ADNPSYNTKTQIIQDFLRKGSAGDGFHGDVYLTVKLLPGVIKTV NLNDKQIVKLFSTRIFNCNPDDMARDLEQGDVSETIRVFFEQSKSFP PAAKSLLTIQEVDEFLLRLSKLTKEDEQQQALQDIASRCTANDLKCI IRLIKHDLKMNSGAKHVLDALDPNAYEAFKASRNLDQVVVERVLHNA QEVEKEPGQRRALSQASLMTVPQPMLEACKSVEYAMKKCPN GMFSEIKYDGERVQVHKNGDHFSYFSRSLKPVLPHKVAHFKDYIP QAFPGGHSMLDSEVLLIDNKTGKPLPFGTLGVHKKAAFQDANVCL FVFDCIYFNDVSLMDRPLCERRKFLHDNMVEIPNRMFSEMKRVTK ALDLADMITRVIQEGLEGLVLKDVKGTYEPGKRHWLKVKKDYLNE GAMADTADLVVLGAFYGGQSGKGMMSIFLMGCYDPGSQKWCTV TKCAGGHDDATLARLQNELDMVKISKDPSKIPSWLVNKIYYPDFI VPDPKKAHVWEITGAEFSKSEAHTADGISIRFPRCTRIRDDKDWKS ATNLPQLKELYQLSKEADFTVVAGDEGSSTTGGSSSEENKGPSG SAVSRKAPSKPSASTKKAEGKLSNSNSKDGNMQTAKPSAMKVGE KLATKSSPVKVGKRAADETLCQTKVLLDIFTGVRLYLPPSTPDF SRLRRYFVAFDGLVQEFDMTSATHVLGSRDKNPAAQQVSPewi WACIRKRRLVAPC
12	Full amino acid sequence of human DNA polymerase beta (PB) (M at start of protein sequence is removed to fuse to Cas9)	MSKRKAPQETLNGGITDMLTELANFEKNVSQAIHKYNAYRKAASVI AKYPHKKISGAEAKKLPVGTGKIAEKIDEFLATGKLRLKLEKIRQDDT SSSINFLTRVSGIGPSAARKFVDEGIKTLEDLRKNEDKLNHHQRIGL KYFGDFEKRIPREEMLQMMDIVLNEVKKVDSEYIATVCGSFRRGA ESSGDMDVLLTHPSFTSESTKQPKLLHQVVEQLQKVHFITDTLSK GETKFMGVCQLPSKNDEKEYPHRRIDIRLIPKDQYYCGVLYFTGS DIFNKNMRAHALEKGFTINEYTIRPLGVTGVAGEPLPVDSEKDIFDY IQWKYREPKDRSE
13	Full amino acid sequence of rat DNA polymerase beta (PB) (M at start of protein)	MSKRKAPQETLNGGITDMLVELANFEKNVSQAIHKYNAYRKAASVI AKYPHKKISGAEAKKLPVGTGKIAEKIDEFLATGKLRLKLEKIRQDDT SSSINFLTRVTGIGPSAARKLVDEGIKTLEDLRKNEDKLNHHQRIGL KYFEDFEKRIPREEMLQMMDIVLNEVKKLDPEYIATVCGSFRRGAE SSGDMDVLLTHPNFTSESSKQPKLLHRVVEQLQKVRFITDTLSKG ETKFMGVCQLPSENDENEYPHRRIDIRLIPKDQYYCGVLYFTGSDI FNKNMRAHALEKGFTINEYTIRPLGVTGVAGEPLPVDSEQDIFDYI QWRYREPKDRSE

	sequence is removed to fuse to Cas9)	
14	EMX1 gRNA sequence (spacer is underlined)	<u>GAGTCCGAGCAGAAGAAG</u> AGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
15	FANCF gRNA sequence (spacer is underlined)	<u>GGAATCCCTTCTGCAGCACCG</u> TTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
16	HEK2-1 gRNA sequence (spacer is underlined)	<u>GAACACAAAGCATAGACTGCG</u> TTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
17	HEK2-2 gRNA sequence (spacer is underlined)	<u>GAGCACCAACCCCTAACTAG</u> TTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
18	HEK2-3 gRNA sequence (spacer is underlined)	<u>GGAAACGGATAGTTCTGAAAG</u> TTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
19	HEK2-4 gRNA sequence (spacer is underlined)	<u>CTTAAGTATTTGTATTCCACG</u> TTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
20	HEK2-5 gRNA sequence (spacer is underlined)	<u>CTTCCCAAGTGAGAAGCCAG</u> TTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
21	HEK2-6 gRNA sequence (spacer is underlined)	<u>CCAGCCCGCTGGCCCTGTAAG</u> TTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
22	HEK2-7 gRNA sequence (spacer is underlined)	<u>CATTCCGTTATTTTACATATG</u> TTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC

23	HEK2-8 gRNA sequence (spacer is underlined)	<u>GTTTCCTTTACAGGGCCAGCG</u> TTTTAGAGCTAGAAATAGCAAGT TAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGA GTCGGTGC
24	HEK2-9 gRNA sequence (spacer is underlined)	<u>ATACGCACAGTTTGACAGAT</u> GTTTTAGAGCTAGAAATAGCAAGT TAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGA GTCGGTGC
25	HEK2-10 gRNA sequence (spacer is underlined)	<u>GCTGGCCCTGTAAAGGAAAC</u> GTTTTAGAGCTAGAAATAGCAAG TAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCG AGTCGGTGC
26	HEK2-11 gRNA sequence (spacer is underlined)	<u>GCATGCGTGTGTGTTTAAGC</u> GTTTTAGAGCTAGAAATAGCAAGT TAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGA GTCGGTGC
27	HEK2-12 gRNA sequence (spacer is underlined)	<u>TTGGGCTGCAGTAACTTGAAG</u> TTTTAGAGCTAGAAATAGCAAGT TAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGA GTCGGTGC
28	HEK2-13 gRNA sequence (spacer is underlined)	<u>TCTTCAAGCAGGTGATTACG</u> TTTTAGAGCTAGAAATAGCAAGT TAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGA GTCGGTGC
29	HEK2-14 gRNA sequence (spacer is underlined)	<u>AGTTTCCTTTACAGGGCCAGG</u> TTTTAGAGCTAGAAATAGCAAGT TAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGA GTCGGTGC
30	HEK2-15 gRNA sequence (spacer is underlined)	<u>GAGGTCGTGGCTGAGCACA</u> AGTTTTAGAGCTAGAAATAGCAAG TAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCG AGTCGGTGC
31	HEK2-16 gRNA sequence (spacer is underlined)	<u>GGCCTCTATTGTTGGTAGAA</u> GTTTTAGAGCTAGAAATAGCAAGT TAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGA GTCGGTGC
32	HEK3 gRNA sequence	<u>GGCCCAGACTGAGCACGTGAG</u> TTTTAGAGCTAGAAATAGCAAG TAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCG AGTCGGTGC

	(spacer is underlined)	
33	HEK4 gRNA sequence (spacer is underlined)	<u>GGCACTGCGGCTGGAGGTGGGTTTTAGAGCTAGAAATAGCAA</u> GTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCC GAGTCGGTGC
34	RNF2-1 gRNA sequence (spacer is underlined)	<u>GTCATCTTAGTCATTACCTGGTTTTAGAGCTAGAAATAGCAAGT</u> TAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCCGA GTCGGTGC
35	RNF2-2 gRNA sequence (spacer is underlined)	<u>CACACACACTTAGAATCTGTGTTTTAGAGCTAGAAATAGCAAGT</u> TAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCCGA GTCGGTGC
36	RNF2-3 gRNA sequence (spacer is underlined)	<u>ACACACACACTTAGAATCTGGTTTTAGAGCTAGAAATAGCAAGT</u> TAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCCGA GTCGGTGC
37	GJB2 gRNA sequence (spacer is underlined)	<u>GGACACGAAGATCAGCTGCAGTTTTAGAGCTAGAAATAGCAAG</u> TTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCCG AGTCGGTGC
38	ADRB2 gRNA sequence (spacer is underlined)	<u>CCCTTTCCTGCGTGACGTCGGTTTTAGAGCTAGAAATAGCAAG</u> TTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCCG AGTCGGTGC
39	MYBPC3 gRNA sequence (spacer is underlined)	<u>CCCTTTCCTGCGTGACGTCGGTTTTAGAGCTAGAAATAGCAAG</u> TTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCCG AGTCGGTGC
40	GAL 292 gRNA sequence (spacer is underlined)	<u>GAAGTCGTTGTCAAACAGGAGTTTTAGAGCTAGAAATAGCAAGT</u> TAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCCGA GTCGGTGC
41	VEGFA-1 gRNA sequence (spacer is underlined)	<u>GATGTCTGCAGGCCAGATGAGTTTTAGAGCTAGAAATAGCAAG</u> TTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCCG AGTCGGTGC

[00102] In addition to the table above, full amino acid sequences of specific fusion proteins (CGBEs) as disclosed herein are provided in Table 4 below:

SEQ ID NO:	Name	Sequence
42	Amino acid sequence of ACX, hPB(8kD) (Apobec is in bold; nCas9 is in parentheses []; hPB(8kD) is underlined)	MSSETGPVAVDPTLRRRIEPHEFEVFFDPREL RKETCLLYEINWG GRHSIWRHTSQNTNKHVEVNFIEKFTTERTYFCPNTRCSITWFLSW SPCGECSRAITEFLSRYPHVT LFIYIARLYHHADPRNRQGLRDLIS SGVTIQIMTEQESGYCWRNFVNYSNPSNEAHWPRYPHLWVRLYVL ELYCIILGLPPCLNLRKQPKLTFFTIALQSCHYQRLPPHILWATG LKSGSETPGTSESATPES [DKKYSIGLAIGTNSVGWAVITDEYKVP SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTR RKNRICYLQEIFSNEMAKVDDSFHRL EESFLVEEDKKHERHPIFG NIVDEVAYHEKYPTIYHLRKKLV DSTDKADRLIYLALAHMIKFRGH FLIEGDLNPDNSDVKLFIQLVQTYNQ LFEENPINASGVDAKAILS RLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAED AKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRV NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRK QRTFDNGSIPHQIHLGELHAILRRQEDFYFPFLKDNREKIEKILTFRIP YYVGPLARGNSRF AWMTRKSEETITPWNFEVVDKGASAQSFIER MTNFDKNLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAF LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDR FNASLGTYHDL LKIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTI LDFLKSDGFANRNF MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA NLAGSPAIKKGILQTVKVVD ELVKVMGRHKPENIVIAMARENQTTQ KGQKNSRERMKRIEEG IKELGSQILKEHPVENTQLQNEKLYLYLQ NGRDMYVDQELDINRLSDYDV DHIVPQSFLKDDSIDNKVLTRSDK NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG GLSELDKAGFIKRLVETRQITKHVAQILDSRMNTKYDENDKLIREV KVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG TALIKK YPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFF KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQV NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK KYGGFDSPT VAYSVLVVAKVEKGSKKLKSVKELLGITIMERSSEFKNPIDFLEAK GYKEVKKDLI IKLPKYSLFEL ENGRKRMLASAGELQKGNELALPSK YVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII EQISEFS KRVILADANLDKVL SAYNKH RDKPIREQAENIIHLFTLTNLGAPAAF KYFDTTIDRKRYTSTKEVLDATLIHQ SITGLYETRIDLSQLGGD]SGG SSKRKAPQETLNGGITDMLTEL ANFEKNVSQAIHKYNAYRKAASVI <u>AKYPHKIKSGAEAKKLP</u> GVGT KIAEKIDEFLATGKL RKLEKSGGSP KKKRKV
43	Amino acid sequence of ACX, hPB(14kD) (Apobec is in bold; nCas9 is in parentheses []; hPB(14kD) is underlined)	MSSETGPVAVDPTLRRRIEPHEFEVFFDPREL RKETCLLYEINWG GRHSIWRHTSQNTNKHVEVNFIEKFTTERTYFCPNTRCSITWFLSW SPCGECSRAITEFLSRYPHVT LFIYIARLYHHADPRNRQGLRDLIS SGVTIQIMTEQESGYCWRNFVNYSNPSNEAHWPRYPHLWVRLYVL ELYCIILGLPPCLNLRKQPKLTFFTIALQSCHYQRLPPHILWATG LKSGSETPGTSESATPES [DKKYSIGLAIGTNSVGWAVITDEYKVP SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTR RKNRICYLQEIFSNEMAKVDDSFHRL EESFLVEEDKKHERHPIFG NIVDEVAYHEKYPTIYHLRKKLV DSTDKADRLIYLALAHMIKFRGH FLIEGDLNPDNSDVKLFIQLVQTYNQ LFEENPINASGVDAKAILS RLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAED AKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRV

		NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRK QRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIP YYVGPLARGNSRFAMWTRKSEETITPWNFEEVVDKGASAQSFIER MTNFDKNLPNEKVLPHKSHLLYEYFTVYNELTKVKYVTEGMRKPAF LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDR FNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTI LDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA NLAGSPAIAKKGILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQ KGQKNSRERMKRIEELGKELGSQILKEHPVENTQLQNEKLYLYYLQ NGRDMYVDQELDINRLSDYDQDHIVPQSFLKDDSIDNKVLTRSDK NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG GLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREV KVITLKSCLVSDFRKDFQFYKVRINNYHHAHDAYLNAVVGTAIIK YPKLESEFVYGDYKVDVRKMIKSEQEIGKATAKYFFYSNIMNFF KTEITLANGEIRKRLIETNGETGEIVWDKGRDFATVRKVLSPQV NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT VAYSVLVAKVEKGSKKLKSVELLGITIMERSSEKPNIDFLEAK GYKEVKDLIIKLPKYSLELENGRKRMLASAGELQKGNELALPSK YVNFYLAHYEKLKSGPEDNEQKQLFVEQHKHYLDEIIEQISEFS KRVLADANLDKVL SAYNKHDKPIREQAENIHLFTLTNLGAPAAF KYFDTTIDRKRYTSTKEVLDATLIHQSTGLYETRIDLSQLGGD]SGG SSKRKAPQETLNGGITDMLTELANFEKNVSQAIHKYNAYRKAASVI <u>AKYPHIIKSGAEAKKLPGVGTIAEKIDEFLATGKLRKLEKIRQDDT</u> <u>SSSINFLTRVSGIGPSAARKFVDEGIKTLEDLRKNEDKLNHHQRIGL</u> <u>KSGGSPKKKRKV</u>
44	Amino acid sequence of ACX, rPB(8kD) (Apobec is in bold; nCas9 is in parenthese s []; rPB(8kD) is underlined)	MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELKTCCLYEINWG GRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSW SPCGECSRAITEFLSRYPHTLFIYIARLYHHADPRNRQGLRDLIS SGVTIQIMTEQESGYCWRNFVNYSNPSNEAHWPYPHLLWVRLYVL ELYCIILGLPPCLNLRKQPKLTFTTIALQSCHYQRLPPHILWATG LKSGSETPGTSESATPES[DKKYSIGLAIGTNSVGWAVITDEYKVP SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTR RKNRICYLQEIFSNEMAKVDDSSFFHRLEESFLVEEDKKHERHPIFG NIVDEVAYHEKYPTIYHLRKKLVDSTDKADRLIYLALAHMIKFRGH FLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILS RLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAED AKLQLSKDITYDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRV NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRK QRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIP YYVGPLARGNSRFAMWTRKSEETITPWNFEEVVDKGASAQSFIER MTNFDKNLPNEKVLPHKSHLLYEYFTVYNELTKVKYVTEGMRKPAF LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDR FNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTI LDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA NLAGSPAIAKKGILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQ KGQKNSRERMKRIEELGKELGSQILKEHPVENTQLQNEKLYLYYLQ NGRDMYVDQELDINRLSDYDQDHIVPQSFLKDDSIDNKVLTRSDK NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG GLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREV KVITLKSCLVSDFRKDFQFYKVRINNYHHAHDAYLNAVVGTAIIK

		YPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFF KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPMPQV NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKGFFDSPT VAYSVLVVAKVEKGSKKLKSVKELLGITIMERSSEFKNPIDFLEAK GYKEVKKDLIIKLPKYSLEFENGRKRMLASAGELQKGNELALPSK YVNFLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIIEQISEFS KRVILADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPAAF KYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDLSQLGGD]SGG SSKRKAPQETLNGGITDMLVELANFEKNVSQAIHKYNAYRKAASVI <u>AKYPHKKISGAEAKKLPVGTGKIAEKIDEFLATGKLRKLEKSGGSP</u> <u>KKKRRKV</u>
45	Amino acid sequence of ACX, rPB(14kD) (Apobec is in bold; nCas9 is in parentheses s []; rPB(14kD) is underlined)	MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELKRETCCLLYEINWG GRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSW SPCGECSRAITEFLSRYPHTLFIYIARLYHHADPRNRQGLRDLIS SGVTIQIMTEQESGYCWRNFVNYSNPSNEAHWPYPHLLWVRLYVL ELYCIILGLPPCLNLRKQPKLTFFLTALQSCHYQRLPPHILWATG LKSGSETPGTSESATPES [DKKYSIGLAIGTNSVGWAVITDEYKVP SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTR RKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFG NIVDEVAYHEKYPTIYHLRKKLV DSTDKADRLIYLALAHMIKFRGH FLIEGDLNPDNSDVKLFQQLVQTYNQLFEEPNINASGVDAKAILS RLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAED AKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRV NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRK QRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIP YYVGPLARGNSRFWMTRKSEETITPWNFEVVDKGASAQSFIER MTNFDKNLPNEKVLPHKSHLLYEYFTVYNELTKVKYVTEGMRKPAF LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDR FNASLGTYHDLLKIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTI LDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDLSLHEHIA NLAGSPAIIKKGILQTVKVVDLVKVMGRHKPENIVIEMARENQTTQ KGQKNSRERMKRIEIGIKELGSQILKEHPVENTQLQNEKLYLYYLQ NGRDMYVDQELDINRLSDYDVHIVPQSFLKDDSIDNKVLTRSDK NRGKSDNVPSEEVVKMKKNYWRQLLNAKLITQRKFDNLTKAERG GLSELDKAGFIKRLVETRQITKHVAQILDSRMNTKYDENDKLIREV KVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG TALIKK YPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFF KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPMPQV NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKGFFDSPT VAYSVLVVAKVEKGSKKLKSVKELLGITIMERSSEFKNPIDFLEAK GYKEVKKDLIIKLPKYSLEFENGRKRMLASAGELQKGNELALPSK YVNFLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIIEQISEFS KRVILADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPAAF KYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDLSQLGGD]SGG <u>SSKRKAPQETLNGGITDMLVELANFEKNVSQAIHKYNAYRKAASVI</u> <u>AKYPHKKISGAEAKKLPVGTGKIAEKIDEFLATGKLRKLEKIRQDDT</u> <u>SSSINFLTRVTGIGPSAARKLVDEGIKTLEDLRKNEDKLNHHQRIGL</u> <u>KSGGSPKKKRRKV</u>
46	Amino acid sequence of ACX, hPB(8kD),	MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELKRETCCLLYEINWG GRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSW SPCGECSRAITEFLSRYPHTLFIYIARLYHHADPRNRQGLRDLIS SGVTIQIMTEQESGYCWRNFVNYSNPSNEAHWPYPHLLWVRLYVL

	wtCas9 (Apobec is in bold; wtCas9 is in parenthese s []; hPB(8kD) is underlined)	ELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATG LKSGSETPGTSESATPES[DKKYSIGLDIGTNSVGWAVITDEYKVP SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTR RKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFG NIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGH FLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILS RLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAED AKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRV NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRK QRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIP YYVGPLARGNSRFAMWTRKSEETITPWNFEVVDKGASAQSFIER MTNFDKNLPNEKVLPHSLLEYFTVYNELTKVKYVTEGMRKPAF LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDR FNASLGTYHDLLKIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTI LDLFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDLSLHEHIA NLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQ KGQKNSRERMKRIEEGKELGSQILKEHPVENTQLQNEKLYLYYLO NGRDMYVDQELDINRLSDYDVHIVPQSFLKDDSIDNKVLTRSDK NRGKSDNVPSEEVVKKMKNYWRQLNAKLITQRKFDNLTKAERG GLSELDKAGFIKRLVETRQITKHVAQILDSRMNTKYDENDKLIREV KVITLKSCLVSDFRKDFQFYKVRINNYHHAHDAYLNAVVGTAIIKK YPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFF KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPMPQV NIVKKEVQTTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT VAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSEKNPIDFLEAK GYKEVKKDLIILPKYSLFELENGRKRMLASAGELQKGNELALPSK YVNFYLYASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIEQISEFS KRVILADANLDKVL SAYNKHDKPIREQAENIHLFTLTNLGAPAAF KYFDTTIDRKRYTSTKEVLDTLHQSITGLYETRIDLSQLGGD]SGG SSKRKAPQETLNGGITDMLTELANFEKNVSQAIHKYNAYRKAASVI AKYPHKIKSGAEAKKLPVGVGTIAEKIDEFLATGKLRKLEKSGGSP KKKRKV
47	Amino acid sequence of ACX, hPB(14kD), wtCas9 (Apobec is in bold; wtCas9 is in parenthese s []; hPB(14kD) is underlined)	MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELKRETCLLYEINWG GRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSW SPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLIS SGVTIQIMTEQESGYCWRNFVNYSNPSNEAHWPYPHLLWVRLYVL ELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATG LKSGSETPGTSESATPES[DKKYSIGLDIGTNSVGWAVITDEYKVP SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTR RKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFG NIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGH FLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILS RLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAED AKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRV NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRK QRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIP YYVGPLARGNSRFAMWTRKSEETITPWNFEVVDKGASAQSFIER MTNFDKNLPNEKVLPHSLLEYFTVYNELTKVKYVTEGMRKPAF LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDR FNASLGTYHDLLKIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTI

		LDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGGGDSLHEHIA NLAGSPAIAKKGILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQ KGQKNSRERMKRIEEGKELGSQILKEHPVENTQLQNEKLYLYYLQ NGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDK NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG GLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREV KVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTAIIKK YPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFF KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPMPQV NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT VAYSVLVVAKVEKGSKKLKSVKELLGITIMERSSEFKNPIDFLEAK GYKEVKKDLIIKLPKYSLEFENGRKRMLASAGELQKGNELALPSK YVNFYLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIIEQISEFS KRVILADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPAAF KYFDTTIDRKRYTSTKEVLDTLHQSITGLYETRIDLSQLGGD]SGG SSKRKAPQETLNGGITDMLTELANFEKNVSAIHKYNAYRKAASVI AKYPHKKISGAEAKKLPVGTGKIAEKIDEFLATGKLRLKLEKIRQDDT SSSINFLTRVSGIGPSAARKFVDEGIKTLEDLRKNEKLNHHQRIGL KSGGSPKKKRKV
48	Amino acid sequence of ACX, rPB(8kD), wtCas9 (Apobec is in bold; wtCas9 is in parenthese s []; rPB(8kD) is underlined)	MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELKTCCLLYEINWG GRHSIWRHTSQNTNKHVEVNFIEKFITTERYFCPNTRCSITWFLSW SPCGECSRAITEFLSRYPHVTLFYIARLYHHADPRNRQGLRDLIS SGVTIQIMTEQESGYCWRNFVNYSNEAHWPYPHLLWVRLYVL ELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATG LKSGSETPGTSESATPES[DKKYSIGLDIGTNSVGWAVITDEYKVP SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTR RKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFG NIVDEVAYHEKYPTIYHLRKKLVDSTDKADRLIYLALAHMIKFRGH FLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILS RLSKSRRLLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAED AKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRV NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS KNGYAGYIDGGASQEYFYKFIKPILEKMDGTEELLVKLNREDLLRK QRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIP YYVGPLARGNSRFAMWTRKSEETITPWNFEVVDKGASAQSFIER MTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAF LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDR FNASLGTYHDLLKIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTI LDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGGGDSLHEHIA NLAGSPAIAKKGILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQ KGQKNSRERMKRIEEGKELGSQILKEHPVENTQLQNEKLYLYYLQ NGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDK NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG GLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREV KVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTAIIKK YPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFF KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPMPQV NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT VAYSVLVVAKVEKGSKKLKSVKELLGITIMERSSEFKNPIDFLEAK GYKEVKKDLIIKLPKYSLEFENGRKRMLASAGELQKGNELALPSK YVNFYLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIIEQISEFS KRVILADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPAAF KYFDTTIDRKRYTSTKEVLDTLHQSITGLYETRIDLSQLGGD]SGG

		SSKRKAPQETLNGGITDMLVELANFEKNVSQAIHKYNAYRKAASVI AKYPHKIKSGAEAKKLPGVGTAKIAEKIDEFLATGKLRKLEKSGGSP KKKRKV
49	Amino acid sequence of ACX, rPB(14kD), wtCas9 (Apobec is in bold; wtCas9 is in parenthese s []; rPB(14kD) is underlined)	MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELKETCLLYEINWG GRHSIWRHTSQNTNKHVEVNFIEKFTTERTYFCPNTRCSITWFLSW SPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLIS SGVTIQIMTEQESGYCWRNFVNYSNEAHWPYPHLLWVRLYVL ELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATG LKSGSETPGTSESATPES [DKKYSIGLDIGTNSVGWAVITDEYKVP SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTR RKNRICYLQEIFSNEMAKVDDSSFFHRLEESFLVEEDKKHERHPIFG NIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYALAHMIKFRGH FLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKILSA RLSKSRRLLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAED AKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRV NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRK QRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIP YYVGPLARGNSRFAMWTRKSEETITPWNFEVVDKGASAQSFIER MTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAF LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDR FNASLGTYHDLLKIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTI LDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA NLAGSPAIKKGILQTVKVVDLVKVMGRHKPENIVIAMARENQTTQ KGQKNSRERMKRIEIGIKELGSQILKEHPVENTQLQNEKLYLYLQ NGRDMYVDQELDINRLSDYDVHIVPQSFLKDDSIDNKVLTSDK NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG GLSELDKAGFIKRLVETRQITKHVAQILDSRMNTKYDENDKLIREV KVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTAIIKK YPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFF KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPMPQV NIVKKTQVQTTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT VAYSVLVAKVEKGSKKLKSVKELLGITIMERSSEKPNIDFLEAK GYKEVKKDLIILPKYSLFELENGRKRMLASAGELQKGNELALPSK YVNFYLAHYEKLKSGSPEDNEQKQLFVEQHKHYLDEIIIEQISEFS KRVLADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPAAF KYFDTTIDRKRYTSTKEVLDTLHQSITGLYETRIDLSQLGGD]SGG <u>SSKRKAPQETLNGGITDMLVELANFEKNVSQAIHKYNAYRKAASVI</u> <u>AKYPHKIKSGAEAKKLPGVGTAKIAEKIDEFLATGKLRKLEKIRQDDT</u> <u>SSSINFLTRVTGIGPSAARKLVDEGIKTLEDLRKNEDKLNHHQRIGL</u> <u>KSGGSPKKKRKV</u>
50	Amino acid sequence of AXC, hPB(8kD) (Apobec is in bold; nCas9 is in parenthese s []; hPB(8kD) is underlined)	MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELKETCLLYEINWG GRHSIWRHTSQNTNKHVEVNFIEKFTTERTYFCPNTRCSITWFLSW SPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLIS SGVTIQIMTEQESGYCWRNFVNYSNEAHWPYPHLLWVRLYVL ELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATG LKSGSSKRKAPQETLNGGITDMLTELANFEKNVSQAIHKYNAYR KAASVIAKYPHKIKSGAEAKKLPGVGTAKIAEKIDEFLATGKLRKLEK SGSETPGTSESATPES [DKKYSIGLAIGTNSVGWAVITDEYKVPSK KFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRK NRICYLQEIFSNEMAKVDDSSFFHRLEESFLVEEDKKHERHPIFGNIV DEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYALAHMIKFRGHFLIE GDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKILSARLS

		KSRRLLENLIAQLPGEKKNGLFGNLIASLGLTPNFKSNFDLAEDAKL QLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILSDILRVNTEI TKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNG YAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRT FDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIPYYV GPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASQSFIERMT NFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFLS GEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRF NASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEER LKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTIL DFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA NLAGSPAIIKKGILQTVKVVDLVKVMGRHKPENIVIAMARENQTTQ KGQKNSRERMKRIEEGKELGSQILKEHPVENTQLQNEKLYLYLQ NGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDK NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG GLSELDKAGFIKRLVETRQITKHVAQILDSRMNTKYDENDKLIREV KVITLKSCLVSDFRKDFQFYKVRINNYHHAHDAYLNAVVGTAIIKK YPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFF KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPMPQV NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT VAYSVLVAKVEKGSKKLKSVKELLGITIMERSSEFKNPIDFLEAK GYKEVKKDLIIKLPKYSLELENGRKRMLASAGELQKGNELALPSK YVNFYLASHYEKLKGPEDNEQKQLFVEQHKHYLDEIIQISEFS KRVLADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPAAF KYFDTTIDRKRYTSTKEVLDTLHQSITGLYETRIDLSQLGGDJS GG SPKKKRKV
51	Amino acid sequence of AXC, hPB(14kD) (Apobec is in bold; nCas9 is in parenthese s []; hPB(14kD) is underlined)	MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELKTCCLLYEINWG GRHSIWRHTSQNTNKHVEVNFIEKFITTERYFCPNTRCSITWFLSW SPCGECSRAITEFLSRYPHVTLFIIYIARLYHHADPRNRQGLRDLIS SGVTIQIMTEQESGYCWRNFVNYSNPSNEAHWPYPHLLWVRLYVL ELYCIILGLPPCLNLRKQPKLTFFTIALQSCHYQRLPPHILWATG <u>LKSGGSSKRKAPQETLNGGITDMLTELANFEKNVSQLHKYNAYR</u> <u>KAASVIAKYPHKIKSGAEAKKLPVGTGKIAEKIDEFLATGKLRKLEK</u> <u>RQDDTSSSINFLTRVSGIGPSAARKEVDEGIKTLEDLRKNEDKLNH</u> <u>HQRIGLKSGETPGTSESATPES[DKKYSIGLAIGTNSVGWAVITDE</u> YKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR RYTRRKNRICYLQEIFSNEMAKVDDSFHRLVESFLVEEDKKHER HPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYALAHMIK FRGHFLIEGDLNPDNSDVDFIQLVQTYNQLFEENPINASGVDAK AILSARLSKSRRLLENLIAQLPGEKKNGLFGNLIASLGLTPNFKSNF DLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE DLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKI LTRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASA QSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGM RKPFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIS GVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDR EMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQ SGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDSL HEHIANLAGSPAIIKKGILQTVKVVDLVKVMGRHKPENIVIAMAREN QTTQKGQKNSRERMKRIEEGKELGSQILKEHPVENTQLQNEKLYL YYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTR SDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKA

		ERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL IREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG ALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNI MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGG FDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSEFEKNPID FLEAKGYKEVKKDLIIKLPKYSLEFENGRKRMLASAGELQKGNEL ALPSKYVNFLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIEQI SEFSKRVILADANLDKVL SAYNKHRDKPIREQAENIIHLFTLTNLGA PAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGG DJSGGSPKKKRKV
52	Amino acid sequence of AXC, rPB(8kD) (Apobec is in bold; nCas9 is in parentheses []; rPB(8kD) is underlined)	MSSETGPVAVDPTLRRRIEPHEFEVFFDPREL RKETCLLYEINWG GRHSIWRHTSQNTNKHVEVNFIEKF TTERYFCPNTRCSITWFLSW SPCGECSRAITEFLSRYPHVT LFIYIARLYHHADPRNRQGLRDLIS SGVTIQIMTEQESGYCWRNFVNYS PSNEAHWP RPYPHLWVRLYVL ELYCIILGLPPCLNLRKQPKLTFF TIALQSCHYQRLPPHILWATG LKSGGSSKRKAPQETLNGGITDMLVELANFEKNVSQAIHKYNAYR <u>KAASVIAKYPHKIKSGAEAKKLP GVGTKIAEKIDEFLATGKLRKLEK</u> <u>SGSETPGTSESATPES</u>[DKKYSIGLAIGTNSVGWAVITDEYKVP SK KFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRK NRICYLQEIFS NEMAKVDDSFHRL EESFLVEEDKKHERHPIFGNIV DEVAYHEKYPTIYHLRKKLV DSTDKADLR LIYALAHMIKFRGHFLIE GDLNPDNSDV DKLFIQLVQTYNQLFEENPINASGVDAKILSARLS KSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKL QLSKDTYDDDLNLLAQIGDQYADLFLAAKNLS DAILSDILRVNTEI TKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNG YAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRT FDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTRIPYYV GPLARGNSRFAMTRKSEETITPWNFE EVVDKGASAQSFIERMT NFDKNLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLS GEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRF NASLGTYHDLLKIKDKDFLDNEENEDILEDIVLT LTFEDREMIEER LKTAYHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTIL DFLKSDGFANRNF MQLIHDDSLTFKEDIQKAQVSGGQDSLHEHIA NLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQ KGQKNSRERMKRIE EGIKELGSQILKEHPVENTQLQNEKLYLYYLQ NGRDMYVDQELDINRLSDYDVHIVPQSFLKDDSIDNKVLTRSDK NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG GLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREV KVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG TALIKK YPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFF KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQV NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT VAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSEFEKNPIDFLEAK GYKEVKKDLIIKLPKYSLEFENGRKRMLASAGELQKGNELALPSK YVNFLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIEQISEFS KRVILADANLDKVL SAYNKHRDKPIREQAENIIHLFTLTNLGAPAAF KYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGGD]SGG SPKKKRKV
53	Amino acid sequence of AXC, rPB(14kD) (Apobec is	MSSETGPVAVDPTLRRRIEPHEFEVFFDPREL RKETCLLYEINWG GRHSIWRHTSQNTNKHVEVNFIEKF TTERYFCPNTRCSITWFLSW SPCGECSRAITEFLSRYPHVT LFIYIARLYHHADPRNRQGLRDLIS SGVTIQIMTEQESGYCWRNFVNYS PSNEAHWP RPYPHLWVRLYVL ELYCIILGLPPCLNLRKQPKLTFF TIALQSCHYQRLPPHILWATG

	in bold; nCas9 is in parenthese s []; rPB(14kD) is underlined)	LKSGGSSKRKAPQETLNGGITDMLVELANFEKNVSQAIHKYNAYR KAASVIAKYPHKIKSGAEAKKLPGVGTKIAEKIDEFLATGKLRKLEKI RQDDTSSSINFLTRVTGIGPSAARKLVDEGIKTLEDLRKNEDKLNH HQRIGLKSGSETPGTSESATPES[DKKYSIGLAIGTNSVGWAVITDE YKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR RYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHER HPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLR LIYLALAHMIK FRGHFLIEGDLNPDNSDV DKLFIQLVQTYNQLFEENPINASGVDAK AILSARLSKSRLENLIAQLPGEKKNGLFGNLIASLSGLTPNFKSNF DLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE DLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFLKDNREKIEKI LTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASA QSFIERMTNFDKNLPNEKVLPHKSLLYEYFTVYNELTKVKYVTEGM RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIS GVEDRFNASLGTYHDLKKIKDKDFLDNEENEDILEDIVLTLTLFEDR EMIEERLKTIAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQ SGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDLSL HEHIANLAGSPAIKKGILQTVKVVDLVKVMGRHKPENIVIEMAREN QTTQKGQKNSRERMKRIIEGKELGSQILKEHPVENTQLQNEKLYL YYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTR SDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKA ERGGLSELDAKAGFIKRLVETRQITKHVAQILDSRMNTKYDENDKL IREVKVITLKSCLVSDFRKDFQFYKVRINNYHHAHDAYLNAVVG ALIKKYPKLESEFVYGDYKVYDVRKMIKSEGEIGKATAKYFFYSNI MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGG FDSPTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSEFKNPID FLEAKGYKEVKKDLIKLPKYSLELENGRKRMLASAGELQKGNEL ALPSKYVNFLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIEQI SEFSKRVLADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGA PAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDLSQLGG DJSGGSPKKKRKV
54	Amino acid sequence of XAC, hPB(8kD) (Apobec is in bold; nCas9 is in parenthese s []; hPB(8kD) is underlined)	MSKRKAPQETLNGGITDMLTELANFEKNVSQAIHKYNAYRKAASVI AKYPH KIKSGAEAKKLPGVGTKIAEKIDEFLATGKLRKLEKSGGSS SETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGR HSIWRHTSQNTNKHVEVNFIEKF TTERYFCPNTRCSITWFLSWSP CGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSG VTIQIMTEQESGYCWRNFVNYSNPSNEAHWP RYPHLWVRLYVLEL YCILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK SGSETPGTSESATPES[DKKYSIGLAIGTNSVGWAVITDEYKVPSK KFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRK NRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIV DEVAYHEKYPTIYHLRKKLV DSTDKADLR LIYLALAHMIKFRGHFLIE GDLNPDNSDV DKLFIQLVQTYNQLFEENPINASGVDAKAILSARLS KSRLENLIAQLPGEKKNGLFGNLIASLSGLTPNFKSNFDLAEDAKL QLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILSDILRVNTEI TKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNG YAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRT FDNGSIPHQIHLGELHAILRRQEDFYFLKDNREKIEKILTFRIPYYV GPLARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMT NFDKNLPNEKVLPHKSLLYEYFTVYNELTKVKYVTEGMRKPAFLS GEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRF

		NASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEER LKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTIL DFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDLSLHEHIA NLAGSPAIIKKGILQTVKVVDLVKVMGRHKPENIVIAMARENQTTQ KGQKNSRERMKRIIEGIGELGSQILKEHPVENTQLQNEKLYLYYLQ NGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDK NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG GLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREV KVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTAIIKK YPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFF KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPMPQV NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT VAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAK GYKEVKKDLIIKLPKYSLEFENGRKRMLASAGELQKGNELALPSK YVNFYLYASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIEQISEFS KRVLADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPAAF KYFDTTIDRKRYSTKEVLDTLHQSITGLYETRIDLSQLGGD]SGG SPKKKRKV
55	Amino acid sequence of XAC, hPB(14kD) (Apobec is in bold; nCas9 is in parenthese s []; hPB(14kD) is underlined)	<u>MSKRKAPQETLNGGITDMLTELANFEKNVSQAIHKYNAYRKAASVI</u> <u>AKYPHKIKSGAEAKKLPVGTGKIAEKIDEFLATGKLRLKEKIRQDDT</u> <u>SSSINFLTRVSGIGPSAARKFVDEGIKTLEDLRKNEDKLNHHQRIGL</u> <u>KSGGSSSETGPVAVDPTLRRRIEPHEFEVFFDPRELKRETCLLYEI</u> NWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWF LSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLR DLISSGVTIQIMTEQESGYCWRNFVNYSNEAHWPYPHLLWVR LYVLELYCIILGLPPCLNLRKQPQLTFFTIALQSCHYQRLPPHIL WATGLKSGSETPGTSESATPES[DKKYSIGLAIGTNSVGWAVITDE YKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR RYTRRKNRICYLQEIFSNEMAKVDDSSFFHRLSEESFLVEEDKKHER HPIFGNIVDEVAYHEKYPTIYHLRKKLVDSDDKADRLIYLALAHMIK FRGHFLIEGDLNPDNSDVKLFQILVQTYNQLFEENPINASGVDAK AILSARLSKSRLENLIAQLPGEKKNLFGNLIASLGLTPNFKSNF DLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE DLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKI LTFRIPYYVGPLARGNSRFWMTRKSEETITPWNFEVVVDKGASA QSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGM RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIS GVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTTLTFEDR EMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQ SGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGQGDLSL HEHIANLAGSPAIIKKGILQTVKVVDLVKVMGRHKPENIVIAMAREN QTTQKGQKNSRERMKRIIEGIGELGSQILKEHPVENTQLQNEKLYL YYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTR SDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKA ERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL IREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTA ALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNI MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSP MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGG FDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPID FLEAKGYKEVKKDLIIKLPKYSLEFENGRKRMLASAGELQKGNEL ALPSKYVNFYLYASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIEQI

		SEFSKRVLADANLDKVL SAYNKH RDKPIREQAENIIHLFTLTNLGA PAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGG D]SGGSPKKKRKV
56	Amino acid sequence of XAC, rPB(8kD) (Apobec is in bold; nCas9 is in parenthese s []; rPB(8kD) is underlined)	<u>MSKRKAPQETLNGGITDMLVELANFEKNVSQAIHKYNAYRKAASVI</u> <u>AKYPHKIKSGAEAKKLPGVGT KIAEKIDEFLATGKLRKLEKSGGSS</u> <u>SETGPVAVDPTLRRRIEPHEFEVFFDPREL RKETCLLYEINWGGR</u> <u>HSIWRHTSQNTNKHVEVNFIEKF TTERYFCPNTRCSITWFLSWSP</u> <u>CGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSG</u> <u>VTIQIMTEQESGYCWRNFVNYSNPSNEAHWP RYPHLWVR LYVLEL</u> <u>YCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK</u> SGSETPGTSESATPES[DKKYSIGLAIGTNSVGWAVITDEYKVPSK KFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRK NRICYLQEIFS NEMAKVDDSFHRL EESFLVEEDKKHERHPIFGNIV DEVAYHEKYPTIYHLRKKLV DSTDKADLR LIYLALAHMIKFRGHFLIE GDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLS KSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKL QLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEI TKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNG YAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRT FDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYV GPLARGNSRFAWMTRKSEETITPWNFEVVDKGASAQSFIERMT NFDKNLPNEKVL PKHSLLEYFTVYNELTKVKYVTEGMRKPAFLS GEQKKAIVDLLFKTRNKVTVKQLKEDYFKKIECFDSVEISGVEDRF NASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEER LKT YAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTIL DFLKSDGFANRNF MQLIHDDSLTFKEDIQKAQVSGQGD SLHEHIA NLAGSPA IKGILQTVKVVD ELVKVMGRHKPENIVIE MARENQTTQ KGQKNSRERMKRIE EGikelGSQILKEHPVENTQLQNEKLYLYYLQ NGRDMYVDQELDINRLSDYDV DHIVPQSFLKDDSIDNKVLTRSDK NRGKSDNVPSEEVVKMKNYWRQLLNAKLITQRKFDNLT KAERG GLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREV KVITLKS KLVSDFRKDFQFYK VREINNYHHAHDAYLNAVVG TALIKK YPKLESEFVYGDYKVYDVRKMI AKSEQEIGKATAKYFFYSNIMNFF KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQV NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK KYGGFDSPT VAYSVLVVAKVEKGKSKKLKSVKELLGITIMERS SFEKNPIDFLEAK GYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSK YVNFLYLASHYEKLKGS PEDNEQKQLFVEQHKHYLDEIIEQISEFS KRVLADANLDKVL SAYNKH RDKPIREQAENIIHLFTLTNLGAPAA FKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGGD]SGG SPKKKRKV
57	Amino acid sequence of XAC, rPB(14kD) (Apobec is in bold; nCas9 is in parenthese s []; rPB(14kD) is underlined)	<u>MSKRKAPQETLNGGITDMLVELANFEKNVSQAIHKYNAYRKAASVI</u> <u>AKYPHKIKSGAEAKKLPGVGT KIAEKIDEFLATGKLRKLEKIRQDDT</u> <u>SSSINFLTRVTGIGPSAARKLVDEGIKTLEDLRKNEDKLNHHQRIGL</u> <u>KSGGSSSETGPVAVDPTLRRRIEPHEFEVFFDPREL RKETCLLYEI</u> <u>NWGGRHSIWRHTSQNTNKHVEVNFIEKF TTERYFCPNTRCSITWF</u> <u>LSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLR</u> <u>DLISSGVTIQIMTEQESGYCWRNFVNYSNPSNEAHWP RYPHLWVR</u> <u>LYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHIL</u> <u>WATGLKSGSETPGTSESATPES[DKKYSIGLAIGTNSVGWAVITDE</u> YKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR RYTRRN RICYLQEIFS NEMAKVDDSFHRL EESFLVEEDKKHER HPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLR LIYLALAHMIK FRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAK

		AILSARLSKSRRLLENLIAQLPGEKKNGLFGNLIASLGLTPNFKSNF DLAEDAKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE DLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKI LTFRIPYYVGPLARGNSRFawMTRKSEETITPWNFEVVVDKGASA QSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGM RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIS GVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTTLTFEDR EMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQ SGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDLSL HEHIANLAGSPAIAKKGILQTVKVVDLVKVMGRHKPENIVIAMAREN QTTQKGQKNSRERMKRIIEGKELGSQILKEHPVENTQLQNEKLYL YYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTR SDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKA ERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL IREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG ALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNI MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGG FDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSEFEKNPID FLEAKGYKEVKKDLIIKLPKYSLEFENGKRMLASAGELQKGNEL ALPSKYVNFYLYASHYEKLKGSPEDENEQKQLFVEQHKHYLDEIEQI SEFSKRVLADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGA PAAFKYFDTTIDRKRYTSTKEVLDTLHQSITGLYETRIDLSQLGG DJSGGSPKKKRKV
58	Amino acid sequence of ACX, hXRCC1 (Apobec is in bold; nCas9 is in parentheses s []; hXRCC1 is underlined)	MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELKTKETCLLYEINWG GRHSIWRHTSQNTNKHVEVNFIEKFITTERYFCPNTRCSITWFLSW SPCGECSRAITEFLSRYPHVTLFYIARLYHHADPRNRQGLRDLIS SGVTIQIMTEQESGYCWRNFVNYSNPSNEAHWPYPHLLWVRLYVL ELYCIILGLPPCLNLRKQPKLTFFTIALQSCHYQRLPPHILWATG LKSGSETPGTSESATPES[DKKYSIGLAIGTNSVGWAVITDEYKVP SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTR RKNRICYLQEIFSNEMAKVDDSSFFHRLEESFLVEEDKKHERHPFIG NIVDEVAYHEKYPTIYHLRKKLVDDSTDKADRLIYLALAHMIKFRGH FLIEGDLNPDNSDVKLFQILVQTYNQLFEENPINASGVDAKAILS RLSKSRRLLENLIAQLPGEKKNGLFGNLIASLGLTPNFKSNFDLAED AKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRE DLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKI LTFRIPYYVGPLARGNSRFawMTRKSEETITPWNFEVVVDKGASA QSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGM RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIS GVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTTLTFEDR EMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTI LDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDLSLHEHIA NLAGSPAIAKKGILQTVKVVDLVKVMGRHKPENIVIAMARENQTTQ KGQKNSRERMKRIIEGKELGSQILKEHPVENTQLQNEKLYL YYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDK NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG GLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREV KVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFF

		KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPMPQV NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKGFFDSPT VAYSVLVVAKVEKGSKKLKSVKELLGITIMERSSEKPNIDFLEAK GYKEVKKDLIIKLPKYSLEFENGRKRMLASAGELQKGNELALPSK YVNFLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIIEQISEFS KRVILADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPAAF KYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDLSQLGGD]SGG SGPEIRLRHVVSQSSQDSTHCAENLLKADTYRKWRAAKAGEKTIS VVLQLEKEEQIHSVDIGNDGSFAFVEVLVGSSAGGAGEQDYEVLLV TSSFMSPSESRSRGSNPNRVRMFGPDKLVRAAAEKRWDRVKIVCS QPYSKDSPFGLSFVRFHSPDPKDEAEAPSQKVTVTKLQGFRVKEE DESANSLRPGALFFSRINKTSPVTASDPAGPSYAAATLQASSAASS ASPVSRAIGSTSKPQESPKGKRKLDLNQEEKKTSPKPPAQLSPSV PKRPKLPAPTRTPATAPVPARAQGAVTGKPRGEGTEPRRPRAGP EELGKILQGVVVVLSGFQNPFRSEL RDKALELGAKYRPDWTRDST HLICAFANTPKYSQVLGLGGRIVRKEWVLDCHRMRRRLPSQRYLM AGPGSSSEDEASHSGSGDEAPKLPQKQPQTKTPTQAAGPSS PQKPPTPEETKAASPV LQEDIDIEGVQSEGQDNGAEDSGDTEDEL RRVAEQKEHRLPPGQEEENGEDPYAGSTDENTDSEEHQEPDLPV PELPDFFQGHFFLYGEFPGDERRKLIRYVTA FNGELEDYMSDRV QFVITAEQWDPSEFEEALMDNPSLAFVRPRWIYSCNEKQKLLPHQL YGVVPQASGGSPKKKRKV
59	Amino acid sequence of ACX, hXRCC1, wtCas9 (Apobec is in bold; wtCas9 is in parentheses []; hXRCC1 is underlined)	MSSETGPVAVDPTLRRRIEPHEFEVFFDPREL RKETCLLYEINWG GRHSIWRHTSQNTNKHVEVNFIEKFTTERTYFCPNTRCSITWFLSW SPCGECSRAITEFLSRYPHVT LFIYIARLYHHADPRNRQGLRDLIS SGVTIQIMTEQESGYCWRNFVNYSNPSNEAHWP RYPHLWVRLYVL ELYCIILGLPPCLNLRKQPQLTFFTIALQSCHYQRLPPHILWATG LKSGSETPGTSESATPES[DKKYSIGLDIGTNSVGWAVITDEYKVP SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTR RKNRICYLQEIFSNEMAKVDDSFHRL EESFLVEEDKKHERHPIFG NIVDEVAYHEKYPTIYHLRKKLVDSTDKADRLIYLALAHMIKFRGH FLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAILSA RLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAED AKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRV NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRK QRTFDNGSIPHQIHLGELHAILRRQEDFYFPFLKDNREKIEKILTFRIP YYVGPLARGNSRF AWMTRKSEETITPWNFE EVVDKGASAQSFIER MTNFDKNLPNEKVL PKHSLLYEYFTVYNELTKVKYVTEGMRKPAF LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDR FNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTI LDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA NLAGSPAIKKGILQTVKVVDLVKVMGRHKPENIVIAMARENQTTQ KGQKNSRERMKRIE EGikelGSQILKEHPVENTQLQNEKLYLYYLO NGRDMYVDQELDINRLSDYDVHIVPQSFLKDDSIDNKVLTRSDK NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG GLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREV KVITLKSCLVSDFRKDFQFYKVRINNYHHAHDAYLNAVVG TALIKK YPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFF KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPMPQV NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKGFFDSPT VAYSVLVVAKVEKGSKKLKSVKELLGITIMERSSEKPNIDFLEAK GYKEVKKDLIIKLPKYSLEFENGRKRMLASAGELQKGNELALPSK

		YVNFLYLASHYEKLGKSPEDNEQKQLFVEQHKHYLDEIIEQISEFS KRVLADANLDKVL SAYNKH RD KPIREQAENIIHLFTLTNLGAPAAF KYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGGD]SGG SGPEIRLRHV VSCSSQDSTHCAENLLKADTYRKWRAAKAGEKTIS <u>VVLQLEKEEQIHSVDIGNDGSAFVEVLVGSSAGGAGEQDYEVLLV</u> <u>TSSFMSPSESRS GSNPNRVRMFGPDKLVRAAAEKRWDRVKIVCS</u> <u>QPYSKDS PFGLSFVRFHSPDPKDEAEAPSQKVTVTKLGQFRVKEE</u> <u>DESANSLRPGALFFSRINKTSPVTASDPAGPSYAAATLQASSAASS</u> <u>ASPVSRAIGSTSKPQES PKGKRKLDLNQEEKKTPSKPPAQLSPSV</u> <u>PKRPKL PAPTRTPATAPV PARAQGA VTGKPRGEGTEPRRPRAGP</u> <u>EELGKILQG VVVVLSGFQNPFRSEL RDKALELGAKYRPDWTRDST</u> <u>HLICAFANTPKYSQVLGLGGRIVRKEWVLDCHRMRRRLPSQRYLM</u> <u>AGPGSSSEDEASHSGSGDEAPKLPQKQPQT KTPTQAAGPS</u> <u>PQKPPTPEETKAASPV LQEDIDIEGVQSEGQDNGAEDSGDTEDEL</u> <u>RRVAEQKEHRLPPGQEENGEDPYAGSTDENTDSEEHQEPPDLPV</u> <u>PELPDFFQGHFFLYGEFP GDERRKLIRYVTA FNGELEDYMSDRV</u> <u>QFVITAEWDPSFEEALMDNPSLAFVRPRWIYSCNEKQKLLPHQL</u> <u>YGVVPQASGGSPKKRKV</u>
60	Amino acid sequence of AXC, hXRCC1 (Apobec is in bold; nCas9 is in parenthese s []; hXRCC1 is underlined)	MSSETGPVAVDPTLRRRIEPHEFEVFFDPREL RKETCLLYEINWG GRHSIWRHTSQNTNKHVEVN FIEKFTTERYFCPNTRCSITWFLSW SPCGECSRAITEFLSRYPHTLFIYIARLYHHADPRNRQGLRDLIS SGVTIQIMTEQESGYCWRNFVNYS PSNEAHWP RYPHLWVRLYVL ELYCIILGLPPCLNLRKQ PQLTFFTIALQSCHYQRLPPHILWATG <u>LKSGSGSPEIRLRHV VSCSSQDSTHCAENLLKADTYRKWRAAKA</u> <u>GEKTISVVLQLEKEEQIHSVDIGNDGSAFVEVLVGSSAGGAGEQD</u> <u>YEVLLVTSSFMSPSESRS GSNPNRVRMFGPDKLVRAAAEKRWDR</u> <u>VKIVCSQPYSKDS PFGLSFVRFHSPDPKDEAEAPSQKVTVTKLGQ</u> <u>FRVKEEDESANSLRPGALFFSRINKTSPVTASDPAGPSYAAATLQA</u> <u>SSAASSASPVSRAIGSTSKPQES PKGKRKLDLNQEEKKTPSKPPA</u> <u>QLSPSVPKRPKL PAPTRTPATAPV PARAQGA VTGKPRGEGTEPR</u> <u>RPRAGPEELGKILQG VVVVLSGFQNPFRSEL RDKALELGAKYRPD</u> <u>WTRDSTHLICAFANTPKYSQVLGLGGRIVRKEWVLDCHRMRRRL</u> <u>PSQRYLMAGPGSSSEDEASHSGSGDEAPKLPQKQPQT KTPT</u> <u>QAAGPSSPQKPPTPEETKAASPV LQEDIDIEGVQSEGQDNGAEDS</u> <u>GDTEDELRRVAEQKEHRLPPGQEENGEDPYAGSTDENTDSEEH</u> <u>QEPPDLVPPELPDFFQGHFFLYGEFP GDERRKLIRYVTA FNGE</u> <u>EDYMSDRVQFVITAEWDPSFEEALMDNPSLAFVRPRWIYSCNE</u> <u>KQKLLPHQLYGVVPQASGSETPGTSESATPES</u>[DKKYSIGLAIGN SVGWAVITDEYKVP SKKFKVLGNTDRHSIKKNLIGALLFDSGETAE ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRL EESFL VEEDKKHERHP IFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLR LIYLALAHMIKFRGHFLIEGDLNPDNSD VDKLFIQLVQTYNQLFEEN PINASGVDAKILSARLSKSRRL ENLIAQLPGEKKNGLFGNLIALSL GLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLA AKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVR QQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGT EELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFY PFL KDNREKIEKILTFRIPIYYVGPLARGNSRF AWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVL PKHSLLYEYFTVYNELT KVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFK KIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILED IVLTTLTFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSR KLINGIRDKQSGKTILDFLKSDGFANRNF MQLIHDDSLTFKEDIQKA QVSGQGD SLHEHIANLAGSPA IKKGILQTVKVVD ELVKVMGRHKP

		ENIVIAMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFL KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLI TQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSR MNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHA HDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIG KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG RDFATVRKVLSPMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARK KDWDPKKYGGFDSPTVAYSVLVVAKEVKGSKKLKSVKELLGITIM ERSSFEKNPIDFLEAKGYKEVKDLIIKLPKYSLFELENGRKRMLAS AGELQKGNELALPSKYVNFYLYASHYEKLKGSPEDEQKQLFVEQ HKHYLDEIIIEQISEFSKRVLADANLDKVL SAYNKHRRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYE TRIDLSQLGGDJS GGSPKKKRKV
61	Amino acid sequence of XAC, hXRCC1 (Apobec is in bold; nCas9 is in parentheses []; hXRCC1 is underlined)	<u>MGPEIRLRHVVSQSSQDSTHCAENLLKADTYRKWRAAKAGEKTIS</u> <u>VVLQLEKEEQIHVSVDIGNDGSAFVEVLVGSSAGGAGEQDYEVLLV</u> <u>TSSFMSPSESRSNSPNRVRMFDPDLVRAAAEKWRDRVKIVCS</u> <u>QPYSKDSPFGLSFVRFHSPDPKDEAEAPSQKVTVTKLQGRVKEE</u> <u>DESANSLRPGALFFSRINKTSPVTASDPAGPSYAAATLQASSAASS</u> <u>ASPVSRAGSTSKPQESPKGKRKLDLNQEEKKTSPKPPAQLSPSV</u> <u>PKRPKLPAPTRTPATAPVPARAQGAVTGKPRGEGTEPRRPRAGP</u> <u>EELGKILQGVVVVLSGFQNPFRSELRDKALELGAKYRPDWTRDST</u> <u>HLICAFANTPKYSQVLGLGGRIVRKEWVLDCHMRRLPSQRYLM</u> <u>AGPGSSSEEDASHSGSGDEAPKLPQKQPQTKTQAGPSS</u> <u>PQKPPTPEETKAASPVLQEDIDIEGVQSEGQDNGAEDSGDTEDEL</u> <u>RRVAEQKEHRLPPGQEEGEDPYAGSTDENTDSEEHQEPDLPV</u> <u>PELPDFFQGKHFFLYGEFPGDERRKLIRYVTA FNGELEDYMSDRV</u> <u>QFVITAEQWDPSEFEEALMDNPSLAFVRPRWIYSCNEKQKLLPHQL</u> <u>YGVVPQASGGSSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKE</u> <u>TCLLYEINWGGRRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNT</u> <u>RCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPR</u> <u>NRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSNPSNEAHWPYR</u> <u>PHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQ</u> <u>RLPPHILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVG</u> WAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL KRTARRRYTRRKNRICYLQEFSNEMAKVDDSFHRLSEESFLVEED KKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLRILIYA LAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINAS GVDKAILSARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPN FKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLP KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVK LNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNRE KIEKILTFRIPIYYVGPLARGNSRFAMWTRKSEETITPWNFEVVVK GASQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYV TEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD SVEISGVEDRFNASLGTYHDLLKIKDKDFLDNEENEDILEDIVLTLT LFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI RDKQSGKTILDFLKSDFANRNFMLIHDDSLTFKEDIQKAQVSG QGDSLHEHIANLAGSPAIIKGIQT VKVVDLVKVMGRHKPENIVIE MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSID NKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKY

		DENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYL NAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAK YFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFAT VRKVLSPQVNVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP KKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSE EKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQ KGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYL DEIIQISEFSKRVILADANLDKVL SAYNKHDKPIREQAENIIHLFTL TNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDL S QLGGD JSGGSPKKKRKV
62	Amino acid sequence of ACX, hLIG3 (Apobec is in bold; nCas9 is in parentheses; hLIG3 is underlined)	MSSETGPVAVDPTLRRRIEPHEFEVFFDPREL RKETCLLYEINWG GRHSIWRHTSQNTNKHVEVNFIEKF TTERYFCPNTRCSITWFLSW SPCGECRAITEFLSRYPHVT LFIYIARLYHHADPRNRQGLRDLIS SGVTIQIMTEQESGYCWRNFVNYS PSNEAHWP RPYPHLWVRLYVL ELYCIILGLPPCLNLRKQ PQLTFFTIALQSCHYQRLPPHILWATG LKSGETPGTSESATPES[DKKYSIGLAIGTNSVGWAVITDEYKVP SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTR RKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPFG NIVDEVAYHEKYPTIYHLRKKLV DSTDKADRLIYLALAHMIKFRGH FLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSA RLSKSRRLLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAED AKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRV NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRK QRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIP YYVGPLARGNSRF AWMTRKSEETITPWNFEVVDKGASAQSFIER MTNFDKNLPNEKVLPHKSHLLYEYFTVYNELTKVKYVTEGMRKPAF LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDR FNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTI LDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA NLGSPAIIKKGILQTVKVVDL VKVMGRHKPENIVIAMARENQTTQ KGQKNSRERMKRIE EGIKELGSQILKEHPVENTQLQNEKLYLYLQ NGRDMYVDQELDINRLSDYDV DHIVPQSFLKDDSIDNKVLTRSDK NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG GLSELDKAGFIKRLVETRQITKHVAQILDSRMNTKYDENDKLIREV KVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTAIIKK YPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFF KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQV NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT VAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSEFEKNPIDFLEAK GYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSK YVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIQISEFS KRVILADANLDKVL SAYNKHDKPIREQAENIIHLFTL TNLGAPAA KYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDL S QLGGD JSGGSPKKKRKV <u>SSKAAGTPKKKAVVQAKLTTTGQVTS PVK GASFVTSTNPRKFSGF</u> <u>SAKPNNSGEAPSSPTPKRSLSSSKCDPRHKDCLLREFRKL CAMV</u> <u>ADNPSYNTKTQIIQDFLRKGSAGDGFHGDVYLT VKLLPGVIKTVY</u> <u>NLNDKQIVKLF SRIFNCNPDDMARDLEQGDVSETIRVFFEQSKSFP</u> <u>PAAKSLLTIQEVDEFLLRLSKLTKEDEQQQALQDIASRCTANDLKI</u> <u>IRLIKHDLKMNSGAKHVLDALDPNAYEAFKASRNLDQDVVERVLHNA</u> <u>QEVEKEPGQRRALSVQASLMTVPVQPM LAEACKSVEYAMKKCPN</u> <u>GMFSEIKYDGERVQVHKNGDHF SYFSRSLKPVLP HKVAHF KDYIP</u> <u>QAFPGGHS MILDSEVLLIDNKTGKPLPFGTLGVHKKAAFDANVCL</u>

		<u>FVFDCIYFNDVSLMDRPLCERRKFLHDNMVEIPNRIMFSEMKRVTK</u> <u>ALDLADMITRVIQEGLEGLVLKDVKGTYEPGKRHWLKVKKDYLINE</u> <u>GAMADTADLVVLGAFYGGGSKGGMMSIFLMGCYDPSQKWCTV</u> <u>TKCAGGHDDATLARLQNELDMVKISKDPSKIPSWLVKNKIYYPDFI</u> <u>VPDPKKAADVWEITGAEFKSEAHADGISIRFPRCTRIRDDKDWKS</u> <u>ATNLPQLKELYQLSKEKADFTVVAGDEGSSTTGGSSSEENKGPSG</u> <u>SAVSRKAPSKPSASTKKAEGKLSNSNSKDGNMQTAKPSAMKVGE</u> <u>KLATKSSPVKVGEKRAADETLCQTKVLLDIFTGVRLYLPPSTPDF</u> <u>SRLRRYFVAFDGDVLVQEFDMTSATHVLGSRDKNPAAQQVSPewI</u> <u>WACIRKRRLVAPCSGGSPKKKRKV</u>
63	Amino acid sequence of ACX, hLIG3, wtCas9 (Apobec is in bold; wtCas9 is in parentheses []; hLIG3 is underlined)	MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELKRETCLLYEINWG GRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSW SPCGECSRAITEFLSRYPHVTLFYIARLYHHADPRNRQGLRDLIS SGVTIQIMTEQESGYCWRNFVNYSNPSNEAHWPYPHLLWVRLYVL ELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATG LKSGSETPGTSESATPESIDKKYSIGLDIGTNSVGWAVITDEYKVP SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTR RKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPFG NIVDEVAYHEKYPTIYHLRKKLVDSDDKADRLIYLALAHMIKFRGH FLIEGDLNPDNSDVKLFQVLVQTYNQLFEENPINASGVDAKAILSA RLSKSRRLLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAED AKLQLSKDITYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRV NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRK QRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIP YYVGPLARGNSRFawmTRKSEETITPWNFEVVDKGASAQSFIER MTNFDKNLPNEKVLPHSLLYEYFTVYNELTKVKYVTEGMRKPAF LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDR FNASLGTYHDLKIIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTI LDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA NLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQ KGQKNSRERMKRIEKGELGQILKEHPVENTQLQNEKLYLYLQ NGRDMYVDQELDINRLSDYVDHIVPQSFLKDDSIDNKVLTRSDK NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG GLSELDKAGFIKRLVETRQITKHVAQILDSRMNTKYDENDKLIREV KVITLKSCLVSDFRKDFQFYKVRINNYHHAHDAYLNAVVGTAIIKK YPKLESEFVYGDYKVYDVRKMIKSEGEIGKATAKYFFYSNIMNFF KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQV NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKGFFDSPT VAYSVLVAKVEKGSKKLKSVKELLGITIMERSSEFKNPIDFLEAK GYKEVKKDLIIKLPKYSLELENGRKRMLASAGELQKGNELALPSK YVNFYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIIEQISEFS KRVILADANLDKVL SAYNKHrdKPIREQAENIIHLFTLTNLGAPAAF KYFDTTIDRKRYTSTKEVLDTLIHQSTGLYETRIDLSQLGGDJSGG SSKAAGTPKKKAVVQAKLTTTGQVTSVPVKGASFVSTSNPRKFSGF SAKPNNSGEAPSSPTPKRSLSSSKCDPRHKDCLLREFRKLKAMV ADNPSYNTKTQIIQDFLRKGSAGDGFHGDVYLTVKLLLPGVIKTVY NLNDKQIVKLFsrIFNCNPDDMARDLEQGDVSETIRVFFEQSKSFP PAAKSLLTIQEVDEFLLRLSKLTKEDEQQQALQDIASRCTANDLKI IRLIKHDLMNMSGAKHVLDALDPNAYEAFKASRNLDQVVERVLHNA QEVEKEPGQRRALSVQASLMTVPVQPLAEACKSVEYAMKKCPN GMFSEIKYDGERVQVHKNGDHFYSRSLKPVLPKHVAHFKDYP QAFPGGHSMLDSEVLLIDNKTGKPLPFGTLGVHKKAAAFQDANVCL

		<u>FVFDCIYFNDVSLMDRPLCERRKFLHDNMVEIPNRIMFSEMKRVT</u> <u>ALDLADMITRVIQEGLEGLVLKDVKGTYEPGKRHWLKVKKDYLNE</u> <u>GAMADTADLVVLGAFYGGGSKGGMMSIFLMGCYDPSQKWCTV</u> <u>TKCAGGHDDATLARLQNELDMVKISKDPSKIPSWLKVNKIYYPDFI</u> <u>VPDPKKAADVWEITGAEFKSEAHADGISIRFPRCTRIRDDKDWKS</u> <u>ATNLPQLKELYQLSKEKADFTVVAGDEGSSTTGGSSSEENKGP</u> <u>SAVSRKAPSKPSASTKKAEGKLSNSNSKDGNMQTAKPSAMKVGE</u> <u>KLATKSSPVKVGEKRKAADETLCQTKVLLDIFTGVRLYLPPSTPDF</u> <u>SRLRRYFVAFDGDVLVQEFDMTSATHVLGSRDKNPAAQQVSP</u> <u>EWIWACIRKRRLVAPCSGGSPKKKRV</u>
64	Amino acid sequence of AXC, hLIG3 (Apobec is in bold; nCas9 is in parentheses []; hLIG3 is underlined)	MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELKETCLLYEINWG GRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSW SPCGECSRAITEFLSRYPHVTLFYIARLYHHADPRNRQGLRDLIS SGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVL ELYCIILGLPPCLNLRKQPQLTFFTIALQSCHYQRLPPHILWATG LKSGGSSKAAGTPKKKAVVQAKLTTTGQVTSPVKGASFVTSTNPR KFSGFSAKPNNSGEAPSSPTPKRSLSSSKCDPRHKDCLLREFRKL CAMVADNPSYNTKTQIIQDFLRKGSAGDGFHGDVYLTVKLLPGVI KTVYNLNDKQIVKLFSRIFNCNPDDMARDLEQGDVSETIRVFFEQS KSFPPAKSLLTIQEVDEFLRLSKLTKEDEQQALQDIASRCTAN DLKCIIRLIKHDLKMNSGAKHVLDALDPNAYEAFKASRNLDQVVER VLHNAQEVEKEPGQRRALSVQASLMTPVQPMLAEACKSVEYAMK KCPNGMFSEIKYDGERVQVHKNGDHFSYFSRSLKPVLPHKVAHF KDYIPQAFPGGHSMILDSEVLLIDNKTGKPLPFGTLGVHKKAAFQD ANVCLFVFDCIYFNDVSLMDRPLCERRKFLHDNMVEIPNRIMFSEM KRVTKALDLADMITRVIQEGLEGLVLKDVKGTYEPGKRHWLKVKK DYLNEGAMADTADLVVLGAFYGGGSKGGMMSIFLMGCYDPSQ KWCTVTTKCAGGHDDATLARLQNELDMVKISKDPSKIPSWLKVNKI YYPDFIVVPDPKKAADVWEITGAEFKSEAHADGISIRFPRCTRIRDD KDWKSATNLPQLKELYQLSKEKADFTVVAGDEGSSTTGGSSSEEN KGPSGSAVSRKAPSKPSASTKKAEGKLSNSNSKDGNMQTAKPSA MKVGEKLATKSSPVKVGEKRKAADETLCQTKVLLDIFTGVRLYLPP STPDFSRLRRYFVAFDGDVLVQEFDMTSATHVLGSRDKNPAAQQV SPEWIWACIRKRRLVAPCSGSETPGTSESATPESIDKKYSIGLAIGT NSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESF LVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADL RLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEE NPINASGVDAKILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALS LGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFL AAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALV RQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDG TEELLVKLNREDLLRKQRTFDNGSIPHQIHGELHAILRRQEDFYPF LKDNREKIEKILTRIPYYVGPLARGNSRFAWMTRKSEETITPWNF EEVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSLLYEYFTVYNEL TKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFK KIECFDSVEISGVEDRFNASLGTYHDLKIIKDKDFLDNEENEDILED IVLTTLFEDREMIERLKYAHLFDDKVMKQLKRRRYTGWGRLSR KLINGIRDKQSGKTILDFLKSDGFANRNFQLIHDDSLTFKEDIQKA QVSGQGDSLHEHIANLAGSPAIKKILQTVKVVDELVKVMGRHKP ENIVEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYVDHIVPQSFL KDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKKNYWRQLLNAKLI TQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSR

		MNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHA HDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIG KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG RDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNSDKLIARK KDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIM ERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDEQKQLFVEQ HKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKHDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQSIITGLYE TRIDLSQLGGDJSJGGSPKKKRV
65	Amino acid sequence of XAC, hLIG3 (Apobec is in bold; nCas9 is in parentheses []; hLIG3 is underlined)	<u>MSKAAGTPKKKAVVQAKLTTTGQVTSPVKGASFVTSTNPRKFSGF</u> <u>SAKPNNSEAPSSPTPKRSLSSSKCDPRHKDCLLREFRKLKAMV</u> <u>ADNPSYNTKTQIIQDFLRKGSAGDGFHGDVYLTVKLLPGVIKTVY</u> <u>NLNDKQIVKLFSTRIFNCNPDDMARDLEQGDVSETIRVFFEQSKSFP</u> <u>PAAKSLLTIQEVDEFLLRLSKLTKEDEQQQALQDIASRCTANDLKC</u> <u>IRLIKHDLKMNSGAKHVLDALDPNAYEAFKASRNLDQDVVERVLHNA</u> <u>QEVKEPGQRRALSVQASLMTVPVQPMLEACKSVEYAMKKCPN</u> <u>GMFSEIKYDGERVQVHKNGDHFSYFSRSLKPVLPKHVAHFKDYP</u> <u>QAFPGGHSMLDSEVLLIDNKTGKPLPFGTLGVHKKAAAFQDANVCL</u> <u>FVFDCIYFNDVSLMDRPLCERRKFLHDNMVEIPNRMFSEMKRVTK</u> <u>ALDLADMITRVIQEGLEGLVLKDVKGTYEPGKRHWLKVKKDYLNE</u> <u>GAMADTADLVVLGAFYGGQSGKGGMMSIFLMGCYDPSQKWCTV</u> <u>TKCAGGHDDATLARLQNELDMVKISKDPSKIPSWLVKNKIYYPDFI</u> <u>VPDPKKAHVWEITGAEFKSEAHADGISIRFPRCTRIRDDKDWKS</u> <u>ATNLPQLKELYQLSKEKADFTVVAGDEGSSTTGGSSSEENKGPSG</u> <u>SAVSRKAPSKPSASTKKAEGKLSNSNSKDGNMQTAKPSAMKVGE</u> <u>KLATKSSPVKVGEKRKAADETLCQTKVLLDIFTGVRLYLPPSTPDF</u> <u>SRLRRYFVAFDGDVLQEFDMTSATHVLGSRDKNPAAQQVSPWEI</u> <u>WACIRKRRLVAPCSGGSSSETGPVAVDPTLRRRIEPHEFEVFFDP</u> <u>RELKRETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTER</u> <u>YFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFYIARLY</u> <u>HHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSNE</u> <u>AHWPRYPHLWVRLYVLELYCIILGLPPCLNLRKQQLTFTTIAL</u> <u>QSCHYQRLPPHILWATGLKSGSETPGTSESATPESIDKKYSIGLAI</u> GTNSVGWAVITDEYKVPSSKKFKVLGNTDRHSIKKNLIGALLFDSGE TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSSFFHRLLEE SFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKA DLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFQVLVQTYNQLF EENPINASGVDAKILSARLSKSRRENLIQALPGEKKNGLFGNLIA LSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADL FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKA LVRRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKM DGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDF YPFLKDNREKIEKILTFRIPIYVGPLARGNSRFAWMTRKSEETITP WNFEEVVDKGASQSFIERMTNFDKNLPNEKVLPHKSHLLYEYFTV YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKE DYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIKDKDFLDNEENE DILEDIVLTTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGW GRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKE DIQKAQVSGQGDSLHEHIANLAGSPAIIKQILQTVKVVDELVKVMG RHKPENIVIAMARENQTTQKGQKNSRERMKRIE EGIKELGSQILKE HPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVP QSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKKNYWRQLL NAKLITQRKFDNLTKAERGGLSELKAGFIKRLVETRQITKHVAQI

		LDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINN YHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKVYDVRKMIKSE QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVW DKGRDFATVRKVLSPQVNVKKTEVQTGGFSKESILPKRNSDKLI ARKKDWDPKKYGGFDSPTVAYSVLVVAKEGKSKKLKSVKELL GITIMERSSEKPNIDFLEAKGYKEVKKDLIIKLPKYSLELENGRKR MLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL FVEQHKHYLDEIIIEQISEFSKRVILADANLDKVL SAYNKHDKPIRE QAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSI TGLYETRIDLSQLGGDJS GGSPKKKRKV
66	Amino acid sequence of ACX, rLIG3 (Apobec is in bold; nCas9 is in parenthese s []; rLIG3 is underlined)	MSSETGPVAVDPTLRRRIEPHEFEVFFDPREL RKETCLLYEINWG GRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSW SPCGECSRAITEFLSRYPHVTLFYIARLYHHADPRNRQGLRDLIS SGVTIQIMTEQESGYCWRNFVNYSNPSNEAHWPYPHLLWVRLYVL ELYCIILGLPPCLNLRKQPKLTFFTTIALQSCHYQRLPPHILWATG LKSGETPGTSESATPES[DKKYSIGLAIGTNSVGWAVITDEYKVP SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTR RKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPFG NIVDEVAYHEKYPTIYHLRKKLV DSTDKADRLIYLALAHMIKFRGH FLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILS RLSKSRRLLENIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAED AKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRV NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRK QRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIP YYVGPLARGNSRFWMTRKSEETITPWNFEVVDKGASAQSFIER MTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAF LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDR FNASLGTYHDLKKIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTI LDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA NLAGSPAIIKKGILQTVKVVDL VKVMGRHKPENIVIAMARENQTTQ KGQKNSRERMKRIE EGIKELGSQILKEHPVENTQLQNEKLYLYLQ NGRDMYVDQELDINRLSDYDVHIVPQSFLKDDSIDNKVLTRSDK NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG GLSELDKAGFIKRLVETRQITKHVAQILDSRMNTKYDENDKLIREV KVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTAIIKK YPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFF KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQV NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT VAYSVLVVAKEGKSKKLKSVKELLGITIMERSSEKPNIDFLEAK GYKEVKKDLIIKLPKYSLELENGRKRMLASAGELQKGNELALPSK YVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIIEQISEFS KRVILADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPAA KYFDTTIDRKRYTSTKEVLDATLIHQSI TGLYETRIDLSQLGGDJS GG STLAFKTLFPRNL CALGRKELCLFSEQHHWPAIRQFSQWSETNLL CGCCLLQRRKPVLSFQRGHLRPRATHLISWSGSHVGLCTGPCEM AEQRFCDYAKRG TAGCKCKCEKIVKGVCRIKVV PNPFSSESGG DMKEWYHIKCMFEKLERARATTKKIEDL TELEGWEELEDNEKEQI SQHIADLSSKTAATPKKKATVQAKLT TTTGQVTS PVKGASFITSTNP RKFSGFSAAKPNNSEQDPSSPAPKTSLSASKCDPKHKDCLLREFR KLCAMVAENPSYNTKTQIIHDFLQKGSTGDGFRGDVYLT VKLLLP VIKSVYNLNDKQIVKLSRIFNCNPDDMARDLEQGDVSETIRVFFE QSKSFPPAAKSLTIQEVDAFLHL SKLTKEDEQQQALQDIASRCT

		<u>ANDLKCIIRLIKHDLMNSGAKHVLDALDPNAYEAFKASRNLDQVV</u> <u>ERVLHNEQEVEKDPGRRRALS</u><u>VQASLMTPVQPM</u><u>LAEACKSIEYA</u> <u>MKKCPNGMFSEIKYDGERVQVHKKGDHFSYFSRSLKPVLP</u><u>HKVA</u> <u>HF KDYIPKAFPGGQSMILDSEVLLIDNNTGKPLPFGTLGVH</u><u>KKAAF</u> <u>QDANVCLFVFD</u><u>CIYFNDVSLMDRPLCERRKFLH</u><u>DNMVEIRNRIMF</u> <u>SEMKQVT</u><u>KASDLADMINRVIREGLEGLVLKDV</u><u>KGTYEPGKRHWL</u><u>K</u> <u>VKKDYLNEGAMADTADLVVLGAFYGGGSKGGMMSIFLMGCYDPD</u> <u>SQKWCTVTKCAGGHDDATLARLQKELDMVKISKDPSKIPSWL</u><u>KIN</u> <u>KIYYPDFIVPDPKKA</u><u>AVWEITGA</u><u>EF</u><u>SRSE</u><u>AHTADG</u><u>SIRFPRCTRIR</u> <u>DDKDWKSATNLPQLKELYQLSKDKADFAVVAGDEGSSTT</u><u>GG</u><u>SNG</u> <u>ENEGTAGSTVPRKGPKGPPSKSSASAKKTEQKLN</u><u>DPS</u><u>SRGGEKL</u> <u>AVKSSPVKVGMRKKA</u><u>ADETPGLTKRRRRASRQRGR</u><u>RAMRTGRRS</u> <u>GGSPKKKRKV</u>
67	Amino acid sequence of ACX, rLIG3, wtCas9 (Apobec is in bold; wtCas9 is in parentheses []; rLIG3 is underlined)	MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELKETCLLYEINWG GRHSIWRHTSQNTNKHVEVNFIEKFFTTERYFCPNTRCSITWFLSW SPCGECSRAITEFLSRYPHVTLFYIARLYHHADPRNRQGLRDLIS SGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVL ELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATG LKSGSETPGTSESATPES<u>[DKKYSIGLDIGTNSVGVAVITDEYKVP</u> <u>SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTR</u> <u>RKNRICYLQE</u><u>IFSNEMAKVDD</u><u>SSFFHRLEESFLVEEDKKHERHPIFG</u> <u>NIVDEVAYHEKYPTIYHLRKKLVD</u><u>STDKADLRIYLALAHMIKFRGH</u> <u>FLIEGDLNPDNSD</u><u>VDKLFIQLVQTYNQLFEENPINASGVDAKAILS</u><u>A</u> <u>RLSKSRRL</u><u>ENLIAQLPGEKKNGLFGNLIALSLGLTPNF</u><u>KS</u><u>NFDLAED</u> <u>AKLQLSKD</u><u>TYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRV</u> <u>NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS</u> <u>KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRK</u> <u>QRTFDNGSIPHQIHLGELHAILRRQEDFY</u><u>PFLKDNREKIEKILTFRIP</u> <u>YYVGPLARGNSRF</u><u>AWMTRKSEETITPWNFE</u><u>EVVDKGASAQSFIER</u> <u>MTNFDKNLPNEKVL</u><u>PKHSLLYEYFTVYNELTKVKYVTEGM</u><u>RKPAF</u> <u>LSGEQKKAIVD</u><u>LLFKTNRKVT</u><u>VKQLKEDYFKKIECFDSVEISGVEDR</u> <u>FNASLGT</u><u>YHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMI</u><u>EE</u> <u>RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTI</u> <u>LDFLKSDGFANRNF</u><u>MQLIHDDSLTFKEDIQKAQVSGGGDSLHEHIA</u> <u>NLAGSPA</u><u>IKKGILQTVKVVDELVKVMGRHKPENIV</u><u>EMARE</u><u>NQTTQ</u> <u>KGQKNSRERM</u><u>KRIE</u><u>EGIKELGSQILKEHPVENTQLQNEKLYLYYLQ</u> <u>NGRDMYVDQELDINRLSDYDVHIVPQSFLKDDSIDNKVLTRSDK</u> <u>NRGKSDNP</u><u>SEEVVKKMKNYWRQLLNAKLITQRKFDNLT</u><u>KAERG</u> <u>GLSELDKAGFIK</u><u>RQLVETRQITKHVAQILDSRMNTKYDENDKLIREV</u> <u>KVITLKS</u><u>KLVSDFR</u><u>KDFQFYK</u><u>VREINNYHHA</u><u>DAYLNAVVG</u><u>TALIKK</u> <u>YPKLESEFVYGDYK</u><u>VYDV</u><u>RKMIAKSE</u><u>QEIGKATAKYFFYS</u><u>NIMNFF</u> <u>KTEITLANGEIRK</u><u>RPLIETNGETGEIVWDKGRDFATVRKVL</u><u>SMPQV</u> <u>NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK</u><u>KYGGFDSPT</u> <u>VAYSVLVVAKVEK</u><u>GSKKLKSVKELLGITIMERS</u><u>SFEKNPIDFLEAK</u> <u>GYKEVKKDLIIKLPKYS</u><u>LFEL</u><u>ENGRKRLASAGELQKGNELALPSK</u> <u>YVNFLYLASHYEKLK</u><u>GSPEDNEQKQLFVEQHKHYLDEII</u><u>EQISEFS</u> <u>KRVILADANLDKVL</u><u>SAYNKHRDKPIREQAENIIHLFTLTNLGAPAA</u><u>F</u> <u>KYFDTTIDR</u><u>KRYTSTKEVLDATLIHQ</u><u>SITGLYETRIDL</u><u>SQLGGD</u><u>SGG</u> <u>STLAFKTLFPRNL</u><u>CALGRKELCLFSEQHHWPAIRQFSQWSETNLL</u> <u>CGCCLLQRRKPVLSFQRGHLRPRATHLISW</u><u>SGSHVGLCTGPCEM</u> <u>AEQRF</u><u>CVDYAKRG</u><u>TAGCKKCKEKIVKGVCRIGKVV</u><u>PNPFSESGG</u> <u>DMKEWYHIKCMFEKLERARATTKKIEDL</u><u>TELEGWEELEDNEKEQI</u> <u>SQHIADLSSKTAATPKKKATVQAKLTTTGQVTSPVKGAS</u><u>FITSTNP</u> <u>RKFSGFSAAKPNNSEQDPSSPAPKTSLSASKCDPKHKDCLLREFR</u>

		<u>KLCAMVAENPSYNTKTQIIHDFLQKGSTGDGFRGDVYLTVKLLLP</u> <u>VIKSVYNLNDKQIVKLF</u><u>SRIFNCNPDDMARDLEQGDVSETIRVFFE</u> <u>QSKSFPPAAKSLLTIQEVDAFLHLSKLTKEDEQQQALQDIASRCT</u> <u>ANDLKCIIRLIKHDLKMNSGAKHVLDALDPNAYEAFKASRNLDQDV</u> <u>ERVLHNEQEVEKDPGRRRALS</u><u>VQASLMTPVQPM</u><u>LAEACKSIEYA</u> <u>MKKCPNGMFSEIKYDGERVQVHKKGDHFSYFSRSLKPVLPHKVA</u> <u>HFKDYIPKAFPGGQSMILDSEVLLIDNNTGKPLPFGTLGVHKKAAF</u> <u>QDANVCLFVFDCIYFNDVSLMDRPLCERRKFLHDNMVEIRNRIME</u> <u>SEMKQVTKASDLADMINRVIREGLEGLVLKDVKGTYEPGKRHWLK</u> <u>VKKDYLNEGAMADTADLVVLGAFYGGGSKGGMMSIFLMGCYDPD</u> <u>SQKWCTVTKCAGGHDDATLARLQKELDMVKISKDPSKIPSWLKIN</u> <u>KIYYPDFIVDPKKA</u><u>AVWEITGA</u><u>EF</u><u>SRSEAHTADGISIRFPRCTRIR</u> <u>DDKDWKSATNLPQLKELYQLSKDKADFAVVAGDEGSSTTGG</u><u>SNG</u> <u>ENEGTAGSTVPRKGPKGPPSKSSASAKKTEQKLN</u><u>DPS</u><u>SRGGEKL</u> <u>AVKSSPVKVGMRKA</u><u>ADETPGLTKRRRASRQRGR</u><u>RAMRTGRRS</u> <u>GGSPKKKRKV</u>
68	Amino acid sequence of XAC, rLIG3 (Apobec is in bold; nCas9 is in parentheses []; rLIG3 is underlined)	<u>MTLAFKTLFPRNLCALGRKELCLFSEQHHWPAIRQFSQWSETNLL</u> <u>CGCCLLQRRKPVLSFORGHLRPRATHLISWSGSHVGLCTGPC</u><u>EM</u> <u>AEQRFCDYAKRGTAGCKKCKEKIVKGVC</u><u>CRIGKVVPNPFS</u><u>ESGG</u> <u>DMKEWYHIKCMFEKLERARATTKIEDLTELEGWEELEDNEKEQI</u> <u>SQHIADLSSKTAATPKKKATVQAKLTTTGQVTSPVKGASFITSTNP</u> <u>RKFSGFSAAKPNNSEQDPSSPAPKTSLSASKCDPKHKDCLLREFR</u> <u>KLCAMVAENPSYNTKTQIIHDFLQKGSTGDGFRGDVYLTVKLLLP</u> <u>VIKSVYNLNDKQIVKLF</u><u>SRIFNCNPDDMARDLEQGDVSETIRVFFE</u> <u>QSKSFPPAAKSLLTIQEVDAFLHLSKLTKEDEQQQALQDIASRCT</u> <u>ANDLKCIIRLIKHDLKMNSGAKHVLDALDPNAYEAFKASRNLDQDV</u> <u>ERVLHNEQEVEKDPGRRRALS</u><u>VQASLMTPVQPM</u><u>LAEACKSIEYA</u> <u>MKKCPNGMFSEIKYDGERVQVHKKGDHFSYFSRSLKPVLPHKVA</u> <u>HFKDYIPKAFPGGQSMILDSEVLLIDNNTGKPLPFGTLGVHKKAAF</u> <u>QDANVCLFVFDCIYFNDVSLMDRPLCERRKFLHDNMVEIRNRIME</u> <u>SEMKQVTKASDLADMINRVIREGLEGLVLKDVKGTYEPGKRHWLK</u> <u>VKKDYLNEGAMADTADLVVLGAFYGGGSKGGMMSIFLMGCYDPD</u> <u>SQKWCTVTKCAGGHDDATLARLQKELDMVKISKDPSKIPSWLKIN</u> <u>KIYYPDFIVDPKKA</u><u>AVWEITGA</u><u>EF</u><u>SRSEAHTADGISIRFPRCTRIR</u> <u>DDKDWKSATNLPQLKELYQLSKDKADFAVVAGDEGSSTTGG</u><u>SNG</u> <u>ENEGTAGSTVPRKGPKGPPSKSSASAKKTEQKLN</u><u>DPS</u><u>SRGGEKL</u> <u>AVKSSPVKVGMRKA</u><u>ADETPGLTKRRRASRQRGR</u><u>RAMRTGRRS</u> <u>GGSSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEIN</u> <u>WGGRHSIWRHTSQNTNKHVEVNFIEKF</u><u>TTERYFCPNTRCSITWFL</u> <u>SWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRD</u> <u>LISSGVTIQIMTEQESGYCWRNFVNYS</u><u>PSNEAHWP</u><u>RYPHLWVRL</u> <u>YVLELYCIILGLPPCLNILRRKQPQLTFFTIALQ</u><u>SCHYQRLPPHILW</u> <u>ATGLKSGSETPGTSESATPES</u><u>[DKKYSIGLAIGTNSVGWAVITDEY</u> <u>KVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRR</u> <u>YTRRKNRICYLQEIFS</u><u>NEMAKVDDSF</u><u>HRLEESFLVEEDKKHERHP</u> <u>IFGNIVDEVAYHEKYPTIYHLRKKLV</u><u>STDKADLR</u><u>LIYALAHMIKFR</u> <u>GHFLIEGDLNPDNSDV</u><u>DKLFIQLVQTYNQLFEENPINASGVD</u><u>AKAIL</u> <u>SARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLA</u> <u>EDAKLQLSKD</u><u>TYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDIL</u> <u>RVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFD</u> <u>QSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLL</u> <u>RKQRTFDNGSIPHQIHLGELHAILRRQEDFY</u><u>PFLKDNREKIEKILTF</u> <u>RIPYYVGPLARGNSRFAWMTRKSEETITPWNFE</u><u>EVVDKGASAQSF</u> <u>IERMTNFDKNLPNEKVLPHKSLLYEYFTVYNELTKVKYVTEGMRKP</u>

		AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVE DRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTTLTFEDREMI EERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGK TILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEH IANLAGSPAIIKKGILQTVKVVDELVKVMGRHKPENIVIAMARENQTT QKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYL QNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSD KNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAER GGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIR EVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTAI KKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMN FFKTEITLANGEIRKPLIETNGETGEIVWDKGRDFATVRKVL SMP QVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSEFKNPIDFL EAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELAL PSKYVNFLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIEQIS EFSKRVLADANLDKVL SAYNKHRDKPIREQAENIIHLFTLTNLGAP AAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGGD] SGGSPKKRKV
69	Amino acid sequence of ACX, rXRCC1 (Apobec is in bold; nCas9 is in parenthese s []; rXRCC1 is underlined)	MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELKETCLLYEINWG GRHSIWRHTSQNTNKHVEVNFIKFTTERRYFCPNTRCSITWFLSW SPCGECSRAITEFLSRYPHTLFIYIARLYHHADPRNRQGLRDLIS SGVTIQIMTEQESGYCWRNFVNYSNEAHWPYPHLLWVRLYL ELYCIILGLPPCLNLRKQPLTFFTIALQSCHYQRLPPHILWATG LKSGSETPGTSESATPES[DKKYSIGLAIGTNSVGWAVITDEYKVP SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTR RKNRICYLQEIFS NEMAKVDDSFHRL EESFLVEEDKKHERHPIFG NIVDEVAYHEKYPTIYHLRKKLVDSTDKADRLIYLALAHMIKFRGH FLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILS RLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAED AKLQLSKD TYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRV NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRK QRTFDNGSIPHQIHLGELHAILRRQEDFYFPFLKDNREKIEKILTFRIP YYVGPLARGNSRF AWMTRKSEETITPWNFE EVVDKGASAQSFIER MTNFDKNLPNEKVL PKHSLLYEYFTVYNELTKVKYVTEGMRKPAF LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDR FNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTI LDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA NLAGSPAIIKKGILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQ KGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ NGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDK NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG GLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREV KVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTAI KKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFF KTEITLANGEIRKPLIETNGETGEIVWDKGRDFATVRKVL SMPQV NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSP T VAYSVLVAKVEKGKSKKLKSVKELLGITIMERSSEFKNPIDFLEAK GYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSK YVNFLYLASHYEKLKGSPEDEQKQLFVEQHKHYLDEIIEQISEFS KRVILADANLDKVL SAYNKHRDKPIREQAENIIHLFTLTNLGAPAA FKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGGD] SGG

		<u>SPEISLRHVVSCSSQDSTHRAENLLKADTYRKWRSKAGEKTISV</u> <u>VLQLEKEEQIHSVDIGNDGSFAFVEVLVGSSAGGATAGEQDYEVLL</u> <u>VTSSFMSPSESRSGSNPNRVRIFGPKLVRAAAEKRWDRVKIVCS</u> <u>QPYSKDSPLYGLSFVKFHSPDPKDEAEAPSQKVTVTKLQGFVRVKEE</u> <u>DDSANSLRPGALFFNRINKAASASASDPAGPSYAAATLQASSAAS</u> <u>SALPVPKVGSSSSKLQEPKPKGRKLDLGLEDKPPSKPSAGPAAL</u> <u>KRPKLPVPSRTPAATPASTPAQKAVPGKPRGEGTEPRGARAGPQ</u> <u>ELGKILQGVVVVLSGFQNPFRSELRDKALELGAKYRPDWTDPDSTH</u> <u>LICAFANTPKYSQVLGLGGRIVRKEWVLD CYRMRRRLPSRRYLMA</u> <u>GLGSSSEDEGDSHSESGEDEAPKLPRKRPQPKAKTQAAGPSSPP</u> <u>RPPTPEETKAPSPGPQDNSDTDGEQSEGRDNGAEDSGDTEDEL</u> <u>RRVAKQREQRQPPAPEENGEDPYAGSTDENTDSEAPSEADLPI</u> <u>ELPDFQFGKHFFLYGEFPGDERRKLIRYVTA FNGELEDYMSDRVQ</u> <u>FVITAEQWDPNFEEALMENPSLAFVRPRWIYSCNEKQKLLPHQLY</u> <u>GVVPQASGGSPKKRKV</u>
70	Amino acid sequence of ACX, rXRCC1, wtCas9 (Apobec is in bold; wtCas9 is in parenthese s []; rXRCC1 is underlined)	MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELKTCCLYEINWG GRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSW SPCGECRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLIS SGVTIQIMTEQESGYCWRNRFVNYSPSNEAHWPYPHVLVRLVYL ELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATG LKSGSETPGTSESATPES[DKKYSIGLDIGTNSVGWAVITDEYKVP SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTR RKNRICYLQEIFS NEMAKVDDSFHRLEESFLVEEDKKHERHPFIG NIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGH FLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILS RLSKSRRLLENLIAQLPGEKKNLFGNLIASLGLTPNFKSNFDLAED AKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRV NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQS KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRK QRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIP YYVGPLARGNSRF AWMTRKSEETITPWNFEEVVDKGASQSFIER MTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAF LSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFVDSVEISGVEDR FNASLGTYHDLLKIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEE RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTI LDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA NLAGSPAIKKGILQTVKVVDLVKVMGRHKPENIVIAMARENQTTQ KGQKNSRERMKRIE EGIKELGSQILKEHPVENTQLQNEKLYLYYLO NGRDMYVDQELDINRLSDYDVHIVPQSFLKDDSIDNKVLTRSDK NRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERG GLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREV KVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTA LIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFF KTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQV NIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT VAYSVLVAKVEKGSKKLKSVKELLGITIMERSSEKPNIDFLEAK GYKEVKKDLI IKLPKYSLFELENGRKRMLASAGELQKGNELALPSK YVNFLYLASHYEKLKGS PEDNEQQLFVEQHKHYLDEIIEQISEFS KRVILADANLDKVL SAYNKH RDKPIREQAENIIHLFTLTNLGAPAAF KYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDLSQLGGD]SGG <u>SPEISLRHVVSCSSQDSTHRAENLLKADTYRKWRSKAGEKTISV</u> <u>VLQLEKEEQIHSVDIGNDGSFAFVEVLVGSSAGGATAGEQDYEVLL</u> <u>VTSSFMSPSESRSGSNPNRVRIFGPKLVRAAAEKRWDRVKIVCS</u> <u>QPYSKDSPLYGLSFVKFHSPDPKDEAEAPSQKVTVTKLQGFVRVKEE</u>

		<u>DDSANSLRPGALFFNRINKAASASASDPAGPSYAAATLQASSAAS</u> <u>SALPVPKVGSSSSKLQEPKKGKRKLDLGLEDSPKPSKPSAGPAAL</u> <u>KRPKLPVPSRTPAATPASTPAQKAVPGKPRGEGTEPRGARAGPQ</u> <u>ELGKILQGVVVVLSGFQNPFRSEL RDKALELGAKYRPDWTDPDSTH</u> <u>LICAFANTPKYSQVLGLGGRIVRKEWVLDCYRMRRRLPSRRYLMA</u> <u>GLGSSSEDEGDSHSESGEDEAPKLPRKRPQPKAKTQAAGPSSPP</u> <u>RPPTPEETKAPSPGPQDNSD TDGEQSEGRDNGAEDSGDTEDEL</u> <u>RRVAKQREQRQPPAPEENGEDPYAGST DENTDSEAPSEADLPI</u> <u>ELPDFFQQKHFFLYGEFPGDERRKLIRYVTA FNGELEDYMSDRVQ</u> <u>FVITAEWDPNFEEALMENPSLAFVRPRWIYSCNEKQKLLPHQLY</u> <u>GVVPQASGGSPKKKRKV</u>
71	Amino acid sequence of AXC, rXRCC1 (Apobec is in bold; nCas9 is in parentheses []; rXRCC1 is underlined)	MSSETGPVAVDPTLRRRIEPHEFEVFFDPREL RKETCLLYEINWG GRHSIWRTS QNTNKHVEVN FIEKFTTERYFCPNTRCSITWFLSW SPCGECSRAITEFLSRYPHVT LFIYIARLYHHADPRNRQGLRDLIS SGVTIQIMTEQESGYCWRNFVNYS PSNEAHWPYPHLLWVRLYVL ELYCIILGLPPCLNLRKQPKLTFFTIALQ SCHYQRLPPHILWATG LKSGGSPEISLRHV VSCSSQDSTHRAENLLKADTYRKWRSKAG EKTISVVLQLEKEEQIHSVDIGNDGS AFVEVLVGSSAGGATAGEQD YEVLLVTSSFMSPSESRS GSNPNRVRIFGPKLVRAAAEKRWD RV KIVCSQPYSKDSPYGLSFVKFHSPD KDEAEAPSQKVTVTKLQGF RVKEEDDSANSLRPGALFFNRINKAASASASDPAGPSYAAATLQA SSAASSALPVPKVGSSSSKLQEPKKGKRKLDLGLEDSPKPSKPSA GPAALKRPKLPVPSRTPAATPASTPAQKAVPGKPRGEGTEPRGA RAGPQELGKILQGVVVVLSGFQNPFRSEL RDKALELGAKYRPDWT PDSTHLICAFANTPKYSQVLGLGGRIVRKEWVLDCYRMRRRLPSR RYLMAGLGSSSEDEGDSHSESGEDEAPKLPRKRPQPKAKTQAAG PSSPPRPPTPEETKAPSPGPQDNSD TDGEQSEGRDNGAEDSGD TEDELRRVAKQREQRQPPAPEENGEDPYAGST DENTDSEAPSEA DLPIELPDFFQQKHFFLYGEFPGDERRKLIRYVTA FNGELEDYMS DRVQFVITAEWDPNFEEALMENPSLAFVRPRWIYSCNEKQKLLP HQLYGVVPQASGSETPGTSESATPES[DKKYSIGLAIGTNSVGWA VITDEYKVP SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEAEFVEEDKK HERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLRILIYALA HMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASG VDAIL SARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNF KSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAKNLSD AILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEK YKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKL NREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPLKDNREKI EKILTRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGA SAQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSV EISGVEDRFNASLGT YHDL LKIKDKDFLDNEENEDILEDIVLT LTLF EDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIR DKQSGKTILDFLKS DGFANRNFMLIHDDSLTFKEDIQKAQVSGQ GDSLHEHIANLAGSPA IKKGILQTVKVVDLVKVMGRHKPENIVIEM ARENQTTQKGQKNSRERMKRIE EGIKELGSQILKEHPVENTQLQN EKLYLYYLQNGRDMYVDQELDINRLSDYDV DHIVPQSFLKDDSIDN KVLRSDKNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRKFD NLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYD ENDKLIREVKVITLKS KLVSDFRKDFQFYKVINNYHHAHDAYLN AVVGTALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKY FFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATV

		RKVLSMPQVNIKKTEVQTTGGFSKESILPKRNSDKLIARKKDWDPK KYGGFDSPTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSE KNPIDFLEAKGYKEVKKDLIIKLPKYSLEFENGRKRMLASAGELQK GNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD EIIQISEFSKRVLADANLDKVL SAYNKHDKPIREQAENIIHLFTLT NLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQSTGLYETRIDL SQ LGGDJSGGSPKKKRV
72	Amino acid sequence of XAC, rXRCC1 (Apobec is in bold; nCas9 is in parenthese s []; rXRCC1 is underlined)	<u>MPEISLRHVSCSSQDSTHRAENLLKADTYRKWRSKAGEKTISV</u> <u>VLQLEKEEQIHSDIGNDGSAFVEVLVGSSAGGATAGEQDYEVLL</u> <u>VTSSFMSPSESRSNSPNRVRIFGPKLVRAAAEKRWDRVKIVCS</u> <u>QPYSKDSPLYGLSFVKFHSPDPKDEAEAPSQKVTVTKLQGFVKEE</u> <u>DDSANSLRPGALFFNRINKAASASASDPAGPSYAAATLQASSAAS</u> <u>SALPVPKVGSSSKLQEPKPKGRKLDLGLLEDSKPPSKPSAGPAAL</u> <u>KRPKLPVPSRTPAATPASTPAQKAVPGKPRGEGTEPRGARAGPQ</u> <u>ELGKILQGVVVVLSGFQNPFRSELRDKALELGAKYRPDWTDPDSTH</u> <u>LICAFANTPKYSQVLGLGGRIVRKEWVLD CYRMRRRLPSRRYLMA</u> <u>GLGSSSEDEGDSHSESGEDEAPKLPRKRPQPKAKTQAAGPSSPP</u> <u>RPPTPEETKAPSPGPQDNSD TDGEQSEGRDN GAEDSGDTEDEL</u> <u>RRVAKQREQRQPPAPEENGEDPYAGSTDENTDSEAPSEADLP</u> <u>ELPDFQKGKHFLLYGEFPGDERRKLIRYVTA FNGELEDYMSDRVQ</u> <u>FVITAEWDPNFEEALMENPSLAFVRPRWIYSCNEKQKLLPHQLY</u> <u>GVVPQASGGSSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKET</u> CLLYEINWGGRRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTR CSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRN RQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYP HLWVRLYVLELYCIILGLPPCLNLRKQPKLTFFTIALQSCHYQR LPPHILWATGLKSGSETPGTSESATPES[DKKYSIGLAIGTNSVGW AVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLK RTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLSEESFLVEEDK KHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADRLIYLAL AHMIKFRGHFLIEGDLNPDNSDVKLFQILVQTYNQLFEENPINAS GVDKAILSARLSKSRRLNLIAQLPGEKKNGLFGNLIALSLGLTPN FKSNDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLP KYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVK LNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNRE KIEKILTFRIPIYVGPLARGNSRFWMTRKSEETITPWNFEVVDK GASAQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYV TEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD SVEISGVEDRFNASLGTYHDLLKIKDKDFLDNEENEDILEDIVLTLT LFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGI RDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQKAQVSG QGDSLHEHIANLAGSPAIKKILQTVKVDELVKVMGRHKPENIVIE MARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ NEKLYLYYLQNGRDMYVDQELDINRLSDYDVHIVPQSFLKDDSID NKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKY DENDKLIREVKVITLKSCLVSDFRKDFQFYKVINNYHHAHDAYL NAVVG TALIKKYPKLESEFVYGDYKVYDVRKMAKSEQEIGKATAK YFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFAT VRKVL SMPQVNIKKTEVQTTGGFSKESILPKRNSDKLIARKKDWDP KKYGGFDSPTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSE EKNPIDFLEAKGYKEVKKDLIIKLPKYSLEFENGRKRMLASAGELQ KGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYL

		DEIEQISEFSKRVILADANLDKVL SAYNKHRDKPIREQAENIIHLFTL TNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSI TGLYETRIDL S QLGGD]SGGSPKKKRKV
--	--	---

[00103] The invention illustratively described herein may suitably be practiced in the absence of any element or elements, limitation or limitations, not specifically disclosed herein. Thus, for example, the terms "comprising", "including", "containing", etc. shall be read expansively and without limitation. Additionally, the terms and expressions employed herein have been used as terms of description and not of limitation, and there is no intention in the use of such terms and expressions of excluding any equivalents of the features shown and described or portions thereof, but it is recognized that various modifications are possible within the scope of the invention claimed. Thus, it should be understood that although the present invention has been specifically disclosed by preferred examples and optional features, modification and variation of the inventions embodied therein herein disclosed may be resorted to by those skilled in the art, and that such modifications and variations are considered to be within the scope of this invention.

[00104] As used in this application, the singular form "a," "an," and "the" include plural references unless the context clearly dictates otherwise. For example, the term "a genetic marker" includes a plurality of genetic markers, including mixtures and combinations thereof.

[00105] As used herein, the term "about", in the context of concentrations of components of the formulations, typically means +/- 5% of the stated value, more typically +/- 4% of the stated value, more typically +/- 3% of the stated value, more typically, +/- 2% of the stated value, even more typically +/- 1% of the stated value, and even more typically +/- 0.5% of the stated value.

[00106] Throughout this disclosure, certain examples may be disclosed in a range format. It should be understood that the description in range format is merely for convenience and brevity and should not be construed as an inflexible limitation on the scope of the disclosed ranges. Accordingly, the description of a range should be considered to have specifically disclosed all the possible sub-ranges as well as individual numerical values within that range. For example, description of a range such as from 1 to 6 should be considered to have specifically disclosed sub-ranges such as from 1 to 3, from 1 to 4, from 1 to 5, from 2 to 4, from 2 to 6, from 3 to 6 etc., as well as individual numbers within that range, for example, 1, 2, 3, 4, 5, and 6. This applies regardless of the breadth of the range.

[00107] Certain examples may also be described broadly and generically herein. Each of the narrower species and sub-generic groupings falling within the generic disclosure also form part of the disclosure. This includes the generic description of the examples with a proviso or

negative limitation removing any subject matter from the genus, regardless of whether or not the excised material is specifically recited herein.

[00108] Other examples are within the following claims and non-limiting examples. In addition, where features or aspects of the invention are described in terms of Markush groups, those skilled in the art will recognize that the invention is also thereby described in terms of any individual member or subgroup of members of the Markush group.

EXAMPLES

[00109] The C•G to G•C Base Editor (CGBE) disclosed in the present application leverage the cell's innate base excision repair (BER) pathway. Among the proteins involved in the BER pathway, the inventors of the present invention have extensively characterized three major players, which are DNA polymerase β , DNA ligase III and XRCC1.

[00110] Because the relative orientation of Cas9 fusions may affect the activity of the fusion, several versions of the CGBE candidates with the fused rAPOBEC and BER proteins at different orientations with respect to nCas9 were tested (Figure 1b). HEK293AAV cells were separately treated with each candidate along with gRNAs designed to target genomic sites HEK2 and HEK3, both of which were previously used to characterize base editors. As controls, BE3 and BE4 were used to treat the cells with the same gRNAs. Since no effective means of C:G to G:C base editing was known at the time, and BE3 has been observed to effect this reaction as a byproduct of C:G to T:A editing⁴, BE3 was used as a benchmark for the CGBEs.

[00111] High-throughput sequencing of the HEK2 site revealed that the CGBE candidates were able to edit C:G to both G:C and T:A. Out of the 31 candidates tested, 20 candidates showed an increased level of C:G to G:C editing relative to BE3 at position 6 within the protospacer (Figure 4a). The next closest available C was at position 4, where editing levels were lower (< 10%) (Figure 4b), suggesting a narrower editing window than BE3. In the best performing candidates, up to 29% C:G to G:C editing and 6% C:G to T:A editing were observed for the CGBE candidates, compared with 14% and 36% respectively for BE4, and 16% and 18% respectively for BE3.

[00112] Similarly, sequencing at the HEK3 site revealed editing of C:G to both G:C and T:A. 12 out of 31 candidates gave C:G to G:C editing levels at position 5 that were higher than that achieved by BE3, with the best performing candidates effecting up to 13% C:G to G:C editing compared with 4% for BE3 (Figure 5a). Despite an increase in C:G to T:A editing with CGBEs compared to BE3, a more than twofold increase in C:G to G:C editing with CGBEs led to an increase in the editing ratio of C:G to G:C compared to C:G to T:A. Compared to position 5,

C:G to G:C editing at position 3 and position 4 are significantly lower (Figures 5b and 5c). From this pilot screen, seven of the best performing candidates were shortlisted for a secondary screen (Figure 1c).

[00113] CGBEs were further tested for C:G to G:C editing at four genomic sites known to be amenable to BE3-mediated editing – EMX1, HEK4, RNF2, and FANCF (Figure 6a). With CGBEs, C:G to G:C edits were efficiently induced (17-24%, compared to 8-10% with BE3) as the predominant product (up to 69% purity) at HEK4 and RNF2. At sites EMX1 and FANCF, up to 9% C:G to G:C editing was observed despite this not being the predominant edit, while BE3 effected up to 3% C:G to G:C editing. Across the sites tested, CGBE candidates exhibit higher indel rates than BE3 (Figure 7a). This could be due to the generation of an abasic site – which is essential for BER – may have led to higher indel rates. From this secondary screen, ACX, rXRCC1 and ACX, rPB(8kD) are selected for further characterization.

[00114] It is then determined if target sequence context impacts editing efficiency. In characterizing the two CGBEs with four gRNAs targeting four disease-associated sites – including dyslipidemia-associated gene ADRB2, hearing loss-associated gene GJB2, and hypertrophic cardiomyopathy-associated gene MYBPC3 – it was observed that not only did CGBEs efficiently interrogate disease-associated genes, but they also gave higher levels of C:G to G:C editing at C's immediately following an A/T (Figures 6b, 8 and Table 2). The difference in editing efficiency between gRNAs might be due to the different motifs within which the targeted C is located (for example, ACA at HEK2, CCA at HEK3; targeted C is underlined). To address such potential differences, 16 gRNAs that targeted 16 different HEK2 sites were designed. These gRNA-site combinations collectively cover all possible NCN motifs with targeted C's at position 6 (Figures 2a and 6c). Sequencing revealed that the most readily edited motif was ACA, while the least edited motif was GCC. Up to 10% C:G to G:C editing was observed for ACA, ACC, ACT, GCT, TCA, and TCT. This observation applied to both ACX, rPB(8kD) and ACX, rXRCC1. From these results, it was proposed that the favored DNA motifs are WCW, ACC, and GCT (W is either A or T) (Figure 2b). This understanding is consistent with the observation that C:G to G:C editing is more efficient at HEK2, HEK4, and RNF2, which all had favored DNA motifs of ACA, ACT, and TCT, whereas EMX1, FANCF, and HEK3, with suboptimal DNA motifs of TCC and CCA, are less amenable to C:G to G:C editing. With a preference for WCW, ACC, and GCT, the CGBEs as disclosed herein most efficiently target 6 out of 16 possible motifs.

[00115] A reduction in editing levels at position 4 relative to position 6 of HEK2 was observed (Figure 4). Similarly, editing levels at position 3 and position 4 of HEK3 are lower than that at

position 5 (Figure 5). Both observations suggested that the candidates might not inherit the exact base editing window of BE3. To elucidate the editing window of the CGBEs, a genomic site that has an alternating 5'-W-C-W-C-W-C-3' sequence was chosen such that gRNAs can be designed with C's located either at every odd position or at every even position (Figure 2c). BE3 C:G to T:A editing rates were assessed between positions 1 and 9, wherein higher editing efficiency was found between positions 4 and 8. For both ACX, rPB(8kD) and ACX, rXRCC1, C:G to G:C editing was observed in a nine nucleotide window from positions 2 to 10. However, only at positions 5 and 6 was C:G to G:C editing the predominant and appreciable outcome.

[00116] Recognizing that simply removing UGI can potentially increase C:G to G:C editing at the expense of C:G to T:A editing, the effect of fusing rXRCC1 or rPB(8kD) to rAPOBEC-nCas9 was hence quantified. Across 28 independent treatments using a variety of gRNAs (Figures 3 and 9), BE3 was s to effect on average 4.4% C:G to G:C editing. Removing UGI from BE3 raised mean C:G to G:C editing to 11.5% but increased undesired indel byproducts to 5.0% (Figure 7b). Fusing rPB(8kD) did not provide additional significant effect ($p > 0.05$). However, fusing rXRCC1 further raised mean C:G to G:C editing levels to 15.4% ($p = 0.01$) and decreased undesired indel rates (3.4%), potentially due to an equilibrium shift from abasic site to repaired base. Meanwhile, the major byproduct C:G to T:A editing stayed between 6% and 9%. The results indicate that UGI removal from BE3 promotes C:G to G:C editing but increases byproducts; rXRCC1 fusion further increases C:G to G:C editing and decreases byproducts. Therefore, rAPOBEC-nCas9-rXRCC1 was identified as a preferred CGBE embodiment that effects C:G to G:C editing at $15.4 \pm 7\%$ efficiency in human cells, at a $68 \pm 14\%$ purity, within a three-nucleotide target window and WCW, ACC, and GCT target sequence contexts.

[00117] As with CRISPR-Cas systems, base editors have been reported to exhibit potential DNA and RNA off-target effects. Because CGBEs share the same APOBEC-nCas9 component with BE3, CGBE and BE3 activities were assessed side-by-side on 29 off-target sites using 5 gRNAs. CGBE and BE3 induced $>0.1\%$ C:G to D:H edits at the same 15 positions (D is either A, G or T; H is either A, C, or T). Only at 2 out of these 15 positions did CGBE induce greater off-target editing than BE3; at the remaining 13 sites, CGBE showed lower off-target activity. While reduction of off-target activity can be attributed to lower C:G to T:A editing at off-target sites, C:G to G:C editing at the same off-target sites increased (Figure 10). No site was found where CGBE induced off-target editing but BE3 did not. These results indicate that CGBE exhibits lower off-target activity at the off-target sites identified for Cas9 and for the BE3 architecture. Nevertheless, there are potentially vast sequence-independent DNA and

RNA off-target effects that would require a detailed dissection in further work before direct translation of CGBEs.

[00118] One limitation of BE3 is its low efficiency in some cell types. With BE3, low C:G to T:A editing was observed in H9 stem cells at five genomic sites (with a maximum of 1.2% C:G to T:A editing at HEK4; Figure 20b). The CGBEs exhibited similarly low C:G to G:C editing efficiencies in the H9 stem cells when evaluated side-by-side. The low editing might be due to chromosomal abnormalities and different methylation profiles in stem cells. It is known that genomic DNA tends to be more highly methylated in undifferentiated stem cells. While methylation does not lead to sequence changes, such epigenetic modifications may reduce editing efficiency by inhibiting deaminase activity. APOBEC activity on methylated C is therefore disfavored. Without the C deamination, base editing cannot be initiated with this APOBEC. Since APOBEC is inefficient at deaminating methylated cytidines, further APOBEC engineering, alongside codon optimization, might be needed to enhance the efficiency of CGBEs in stem cells. In contrast, in the eHAP cell line, moderate levels of editing was observed with the CGBE (ACX, rXRCC1) inducing up to 8.5% C:G to G:C editing at the RNF2 and VEGFA sites, while BE3 induced 0.9% C:G to T:A editing (Figure 17b). The eHAP data suggests that the CGBEs may be moderately efficient in some cell types even if different base editing technology is not. In HTB9 cells – a urinary bladder cancer cell line, both BE3 and the CGBEs can efficiently induce the desired mutations at many sites (up to 17% C:G to G:C editing with CGBE and up to 18% C:G to T:A editing with BE3; Figure 14b). ACX, rPB(8kD) appears to consistently outperform ACX, rXRCC1 in HTB9 cells, indicating that different CGBE architecture can be employed for optimal performance according to cell types. It was demonstrated that CGBEs is functional across multiple cell types, and that absolute efficiency is partially dependent on the cell type and state, features shared with previous base editors.

[00119] CGBEs that target cytosine within a specified window and convert it into guanine as a predominant editing product were then developed. In separate works, Liu and Koblan designed CGBE candidates by fusing UDG (UNG), UdgX, and polymerases with rAPOBEC-nCas9 (Figure 11). Both the CGBEs described by Liu and Koblan and the CGBEs disclosed herein induce a C to U change via rAPOBEC and envision this U to be further converted to an abasic site; but the strategy involved and downstream resolution of this abasic site are distinctly different. In Liu and Koblan's work, fusing UNG to rAPOBEC-nCas9 increases the amount of C:G to G:C edits by further increasing the generation of abasic sites (Figure 12). In contrast, the present disclosure focuses on manipulating DNA repair downstream of abasic site creation, reasoning that APOBEC-nCas9 is already able to generate a significant amount of C:G to G:C editing (Figure 3a, BE3 (no UGI)).

[00120] Koblan and Liu's work seeks to induce base excision and perform a translesion polymerization across the targeted abasic site (Figure 12). This proposed mechanism is unlikely to be the case for the CGBEs of the present disclosure. Instead, the mechanism of the presently disclosed CGBEs is that upon creation of an abasic site in the Cas9-induced R-loop, cellular UNG is displaced by APE1, after which XRCC1 recruits various BER components to repair the abasic site independently of the unedited opposite strand, giving rise to guanine as the major product. Subsequent DNA repair converts the G:G mismatch to G:C. Such a hypothesis would be consistent with a) the tight binding of Cas9 on its target strand, which renders that strand less accessible to other enzymes; b) the accessibility of the deaminated strand as a single-strand of the R-loop; c) the detrimental effect of UDG and UdgX on C:G to G:C editing (Figure 11), which suggests that the persistence of an UDG-bound site or abasic site might act against the C:G to G:C reaction; and d) the C:G to G:C editing effected by the CGBEs of the present disclosure with domains that do not have intrinsic polymerase activity but are key drivers of abasic site repair. While further mechanistic studies and development continue, CGBEs expand the growing suite of precise genome-editing tools that include CBEs, ABEs, CGBEs and prime editors (Figure 11), together enabling the precise and efficient engineering of DNA for research, biological interrogation, and disease correction.

EXPERIMENTAL SECTION

[00121] Constructs and molecular cloning

[00122] The BE3 (Addgene plasmid #73021), prime editor 2 (Addgene plasmid #132775), pegRNA-HEK3_CTT_ins (Addgene plasmid #132778) plasmids are used in the present disclosure. The BE3 plasmid is a mammalian expression plasmid with BE3 being driven by a CMV promoter. hXRCC1 (pTXG-hXRCC1) and hLIG3 (pGEX4T-hLIG3) (Addgene plasmid # 52283 and # 81055 respectively) are also used. The mutation R400Q is introduced to hXRCC1 and N628K is introduced to hLIG3 via blunt-end ligation. Briefly, plasmid containing either hXRCC1 or hLIG3 is amplified via PCR using Q5 Hot Start HiFi 2X Master Mix (NEB, M0494). PCR product is then treated with DpnI (NEB, R0176) and T4 Polynucleotide Kinase (NEB, M0201) at 37°C for 30 minutes and inactivated at 65°C for 20 minutes before being ligated using T4 DNA Ligase (NEB, M0202; room temperature for 2 hours). Ligated product is then transformed into chemically competent 5-alpha *Escherichia Coli* (NEB, C2987). rXRCC1, rLIG3, hPBs, and rPBs were obtained as human codon-optimized de novo synthesized gene fragments (Twist Biosciences). All other oligonucleotides used in the study were de novo synthesized (IDT DNA). To fuse BER proteins with rAPOBEC-nCas9, Q5 Hot Start HiFi 2X Master Mix was used to generate Gibson fragments of the BER proteins as Gibson inserts. After checking PCR products on a gel, Gibson insert and vector were incubated with NEBuilder

HiFi DNA Assembly Master Mix (NEB, E2621) for 1 hour at 50°C. The Gibson reaction is then transformed into chemically competent *Escherichia Coli*.

[00123] All assembled plasmids were Sanger sequenced for sequence verification and were prepared using either the PureYield Plasmid Miniprep System (Promega, A1223) or Plasmid Plus Maxi Kit (Qiagen, 12965).

[00124] Cell culture, transfection, and genomic DNA harvest

[00125] HEK293AAV cells (Agilent, 240073) were maintained in DMEM with GlutaMAX and sodium pyruvate (Thermo Fisher, 10569-010) supplemented with 10% HI FBS (Thermo Fisher) at 37°C and 5% CO₂. HTB-9 cells (ATCC, 5637) were maintained in RPMI-1640 with L-glutamine and sodium bicarbonate (Sigma, R8758) supplemented with 10% HI FBS (Thermo Fisher) and 1% MEM Non-Essential Amino Acids Solution (Thermo Fisher, 11140050) at 37°C and 5% CO₂. Both HTB9 and HEK cells were transfected via lipofection. After cells reached ~80% confluency, they were washed with PBS, pH 7.2 (Thermo Fisher, 20012-027) before being treated with TrypLE Express (Thermo Fisher, 12604). 30,000 cells were added to each well of a 48-well plate one day before transfection. For each well, 750 ng of base-editor plasmid, 250 ng of gRNA plasmid, and 20 ng of GFP plasmid were transfected into these cells using Lipofectamine 3000 (Invitrogen, L3000015) according to the manufacturer's protocol. The media were replaced with fresh media 24 hours after transfection. For prime editing, 750 ng of PE plasmid, 250 ng of pegRNA, and 83 ng of sgRNA were used for transfection. 72 hours after transfection, media were removed; cells were washed with 50 µL PBS, pH 7.2, and genomic DNA was extracted using 50 µL of Quick Extract DNA Extract Solution (Lucigen, QE09050) per well according to manufacturer's protocol. All sample sizes indicate biological replicates.

[00126] Jurkat cells (ATCC, TIB-152, Clone E6-1) were maintained in RPMI-1640 with L-glutamine and sodium bicarbonate (Sigma, R8758) supplemented with 10% HI FBS (Thermo Fisher) and 1% MEM Non-Essential Amino Acids Solution (Thermo Fisher, 11140050) at 37°C and 5% CO₂. 200,000 cells were nucleofected with 750 ng of base editor and 250 ng of gRNA expression plasmids using the SE Cell Line 4D-Nucleofector X Kit S (Lonza) and program CL-120 on the 4D X-Unit.

[00127] HepG2 cells were maintained in IMDM (Thermo Fisher, 31980-030) supplemented with 10% FBS (Thermo Fisher) and 1% NEAA (Thermo Fisher, 11140050) at 37°C and 5% CO₂. 200,000 cells were nucleofected with 750 ng of base editor and 250 ng of gRNA expression plasmids using the SF Cell Line 4D-Nucleofector X Kit S (Lonza) and program EH-100 on the 4D X-Unit.

[00128] eHAP cells (Horizon Discovery, C669) were maintained in IMDM (Thermo Fisher, 31980-030) supplemented with 10% FBS at 37°C and 5% CO₂. 200,000 cells were nucleofected with 750 ng of base editor and 250 ng of gRNA expression plasmids using the SE Cell Line 4D-Nucleofector X Kit S (Lonza) and program DS-138 on the 4D X-Unit.

[00129] H9 stem cells (WiCell, WA09) were maintained in mTeSR1 (Stemcell technology, 85850). 200,000 cells were nucleofected with 1500 ng of base editor and 500 ng of gRNA expression plasmids using the P3 Primary Cell kit (Lonza, V4XP-3024) and program hES H9 program on the 4D X-Unit.

[00130] Sequencing of genomic DNA

[00131] Sites of interest were prepared for high-throughput sequencing via two PCR amplifications – the first PCR amplifies the region of interest while the second PCR adds appropriate sequencing barcodes. The first PCR was performed in commonly used methods. Primers for the second PCR are based off Illumina adaptors. Amplicons from the second PCR were then pooled and gel extracted (Promega, A9282) to make the final library, which was quantified via Qubit fluorometer (Thermo Fisher) and sequenced on an Illumina iSeq 100 according to the manufacturer's protocol. The resultant FASTQ files were analyzed using CRISPResso2. All sample sizes indicate biological replicates.

[00132] Statistical analyses were performed on Matlab. Weblogos were created using Weblogo 3.

References

1. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533, 420-424 (2016).
2. Nishida, K. et al. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* 353, aaf8729 (2016).
3. Gaudelli, N. M. et al. Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. *Nature* 551, 464-471 (2017).
4. Komor, A. C. et al. Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. *Sci. Adv.* 3, eaao4774 (2017).
5. Anzalone, A. V. et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 576, 149-157 (2019).

6. Kurt, I. C. et al. CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. *Nat. Biotechnol.* (2020). <https://doi.org/10.1038/s41587-020-0609-x>
7. Zhao, D. et al. New base editors change C to A in bacteria and C to G in mammalian cells. *Nat. Biotechnol.* (2020). <https://doi.org/10.1038/s41587-020-0592-2>
8. Liu, D. R., Koblan, L. W. Cytosine to Guanine Base Editor. *World Intellectual Property Organization* (2018).

Claims

1. A fusion protein or a protein complex comprising a DNA binding protein (DnaBP), a nucleobase modifying protein (NMP), and a Base Excision Repair associated protein (BERAP); wherein the fusion protein or protein complex does not comprise a Uracil binding protein or a catalytically active DNA polymerase.
2. The fusion protein or protein complex according to claim 1, wherein the nucleobase modifying protein (NMP) is a cytosine deaminase domain.
3. The fusion protein or protein complex according to claim 1 or 2, wherein the BERAP is selected from the group consisting of an AP endonuclease, an end processing enzyme, a catalytically inactive DNA polymerase, a lyase domain, a Flap endonuclease, a DNA ligase, and a scaffold protein involved in the BER pathway.
4. The fusion protein or protein complex according to any of claims 1-3, wherein the BERAP is selected from the group consisting of: a DNA ligase III (LIG3), an XRCC1, a DNA binding or lyase domain of DNA Polymerase beta, a DNA binding or lyase domain of DNA Polymerase delta, a DNA binding or lyase domain of DNA Polymerase epsilon, an AP endonuclease (APE1), Proliferating cell nuclear antigen (PCNA), DNA-(apurinic or apyrimidinic site) lyase (APEX), Poly (ADP-ribose) polymerase (PARP), Flap endonuclease 1 (FEN1), and DNA ligase I (LIG1).
5. The fusion protein or protein complex according to any of claims 1-4, wherein the BERAP is an XRCC1.
6. The fusion protein or protein complex according to any of claims 1-4, wherein the BERAP is a DNA binding or lyase domain of DNA Polymerase beta.
7. The fusion protein or protein complex according to claim 6, wherein the DNA binding or lyase domain of DNA Polymerase beta corresponds to a region contained within amino acids 1-140, 1-120, 1-100, or 1-87 of a full DNA Polymerase beta sequence (SEQ ID NO: 12 or 13).
8. The fusion protein or protein complex according to any of claims 1-4, wherein the BERAP is a DNA Ligase III (LIG3).

9. The fusion protein or protein complex according to any of claims 1-8, wherein the DnaBP comprises a CRISPR-associated (Cas) domain.

10. The fusion protein or protein complex according to claim 9, wherein the Cas domain is selected from the group consisting of a Cas3, a Cas9, a xCas9, a HF-Cas9, a Cas9-NG, a SpRY Cas9, a circularly permuted Cas9, a Cas10 and a Cas12 (also known as Cpf1), a Cas14, a CasX, a Cas Φ , and variants thereof.

11. The fusion protein or protein complex according to claim 9 or 10, wherein the Cas domain is a Cas nickase (nCas).

12. The fusion protein or protein complex according to claim 9 or 10, wherein the Cas domain is a nuclease inactive Cas (dCas).

13. The fusion protein or protein complex according to any of claims 2-12, wherein the cytosine deaminase domain is a deaminase domain from an apolipoprotein B mRNA-editing complex (APOBEC) family deaminase.

14. The fusion protein or protein complex according to any of claims 1-13, wherein the DNA binding protein (DnaBP) is a nickase Cas protein; wherein Base Excision Repair associated protein (BERAP) is selected from the group consisting of an XRCC1, a DNA Ligase III (LIG3), and a DNA binding or lyase domain of DNA Polymerase beta; wherein the nucleobase modifying protein (NMP) is a deaminase domain from an apolipoprotein B mRNA-editing complex (APOBEC) family deaminase.

15. The fusion protein according to any of claims 1-14, wherein the orientation of the DnaBP, NMP and BERAP within the fusion protein is selected from the group consisting of: [NMP]-[DnaBP]-[BERAP], [NMP]-[BERAP]-[DnaBP] and [BERAP]-[NMP]-[DnaBP]; wherein each instance of "[]-" comprises an optional linker.

16. The fusion protein according to claim 15, wherein the orientation of the DnaBP, NMP and BERAP within the fusion protein is [NMP]-[DnaBP]-[BERAP].

17. A fusion protein comprising:

- a first amino acid sequence that is at least 80% identical to the amino acid sequence of any one of SEQ ID NOs: 1-2,

- a second amino acid sequence that is at least 80% identical to the amino acid sequence of SEQ ID NO: 3,
- a third amino acid sequence that is at least 80% identical to the amino acid sequence of SEQ ID NO: 4-10.

18. A fusion protein comprising a sequence of any one of SEQ ID NOs: 42 to 72.

19. The fusion protein of claim 18, wherein the sequence is SEQ ID NO: 44 or SEQ ID NO: 69.

20. A protein complex comprising:

- a first protein comprising an amino acid sequence that is at least 80% identical to the amino acid sequence of any one of SEQ ID NOs: 1-2,
- a second protein comprising an amino acid sequence that is at least 80% identical to the amino acid sequence of SEQ ID NO: 3,
- a third protein comprising an amino acid sequence that is at least 80% identical to the amino acid sequence of SEQ ID NO: 4-11.

21. A protein-nucleic acid complex comprising a nucleic acid molecule and any one of:
the fusion protein or protein complex according to any one of claims 1-14,
the fusion protein according to any one of claims 15-19; and
the protein complex of claim 20.

22. A pharmaceutical composition comprising the fusion protein or protein complex of any one of claims 1-14, the fusion protein according to any one of claims 15-19, the protein complex of claim 20, or the protein-nucleic acid complex of claim 21.

23. The pharmaceutical composition of claim 22, further comprising a pharmaceutically acceptable excipient.

24. A method of replacing a cytosine with a guanine on a DNA strand in a cell, said method comprises introducing to the cell the fusion protein or protein complex of any one of claims 1-13, the fusion protein according to any one of claims 15-19, the protein complex of claim 20, or the protein-nucleic acid complex of claim 21.

25. A polynucleotide encoding the fusion protein or protein complex of any one of claims 1-14, the fusion protein according to any one of claims 15-19, the protein complex of claim 20, or the protein-nucleic acid complex of claim 21.

26. A vector comprising the polynucleotide of claim 25.

27. A cell comprising the fusion protein or protein complex of any one of claims 1-14, the fusion protein according to any one of claims 15-19, the protein complex of claim 20, or the protein-nucleic acid complex of claim 21.

28. A method of treating a subject having or suspected of having a disease or disorder comprising administering the fusion protein or protein complex of any one of claims 1-14, the fusion protein according to any one of claims 15-19, the protein complex of claim 20, the protein-nucleic acid complex of claim 21, the pharmaceutical composition of any one of claims 22-23, the polynucleotide of claim 25, or the vector of claim 26 to the subject.

29. The method of claim 28, wherein the disease or disorder comprises one or more C>G and/or one or more G>C mutations.

30. The method of claim 29, wherein the disease or disorder is selected from a group consisting of skin fibrosis, bladder cancer, liver cancer, Myasthenic syndrome, Spondyloepimetaphyseal dysplasia, Parkinson's disease, Deafness, blood disorders, and Schnyder crystalline corneal dystrophy.