Genome Editing and Engineering

Course No: BT-637

LECTURE-25

Dr. Kusum K. Singh
Department of Biosciences and Bioengineering
Indian Institute of Technology Guwahati

Published in final edited form as:
Nature. ; 533(7603): 420–424. doi:10.1038/nature17946.

Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage

Alexis C. Komor[1,2], **Yongjoo B. Kim**[1,2], **Michael S. Packer**[1,2], **John A. Zuris**[1,2], and **David R. Liu**[1,2,*]

[1]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, 02138

[2]Howard Hughes Medical Institute, Harvard University, Cambridge, MA, 02138

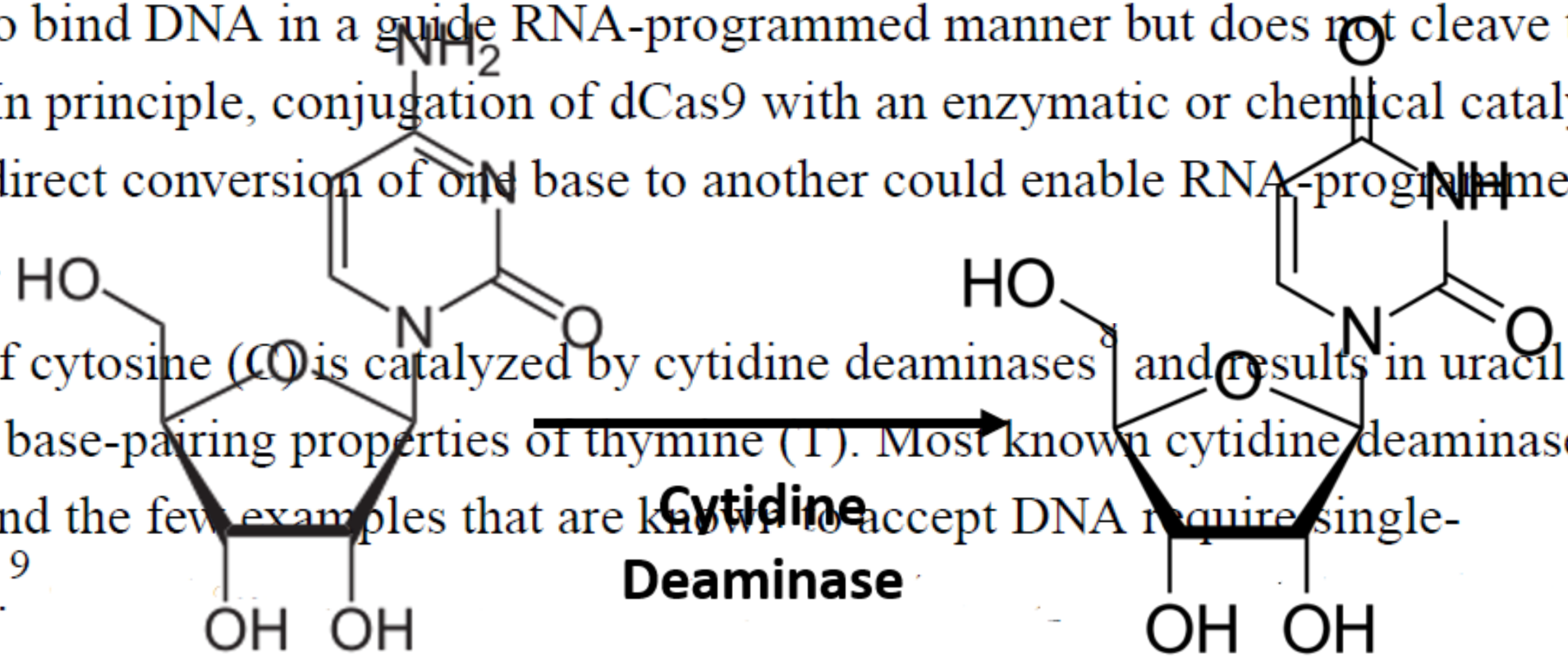
Abstract

Current genome-editing technologies introduce double-stranded (ds) DNA breaks at a target locus as the first step to gene correction.[1,2] Although most genetic diseases arise from point mutations, current approaches to point mutation correction are inefficient and typically induce an abundance of random insertions and deletions (indels) at the target locus from the cellular response to dsDNA breaks.[1,2] Here we report the development of base editing, a new approach to genome editing that enables the direct, irreversible conversion of one target DNA base into another in a programmable

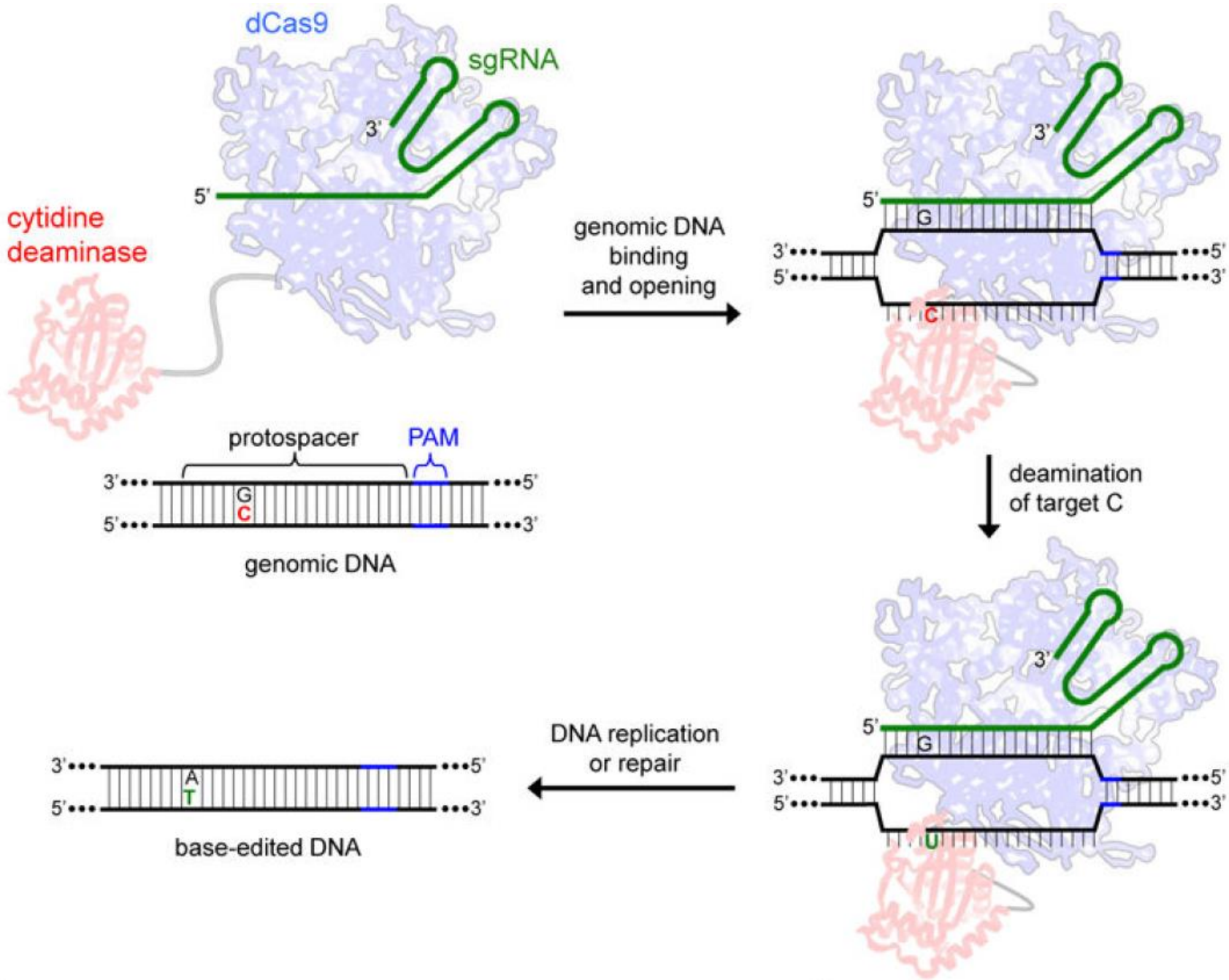
manner, without requiring dsDNA backbone cleavage or a donor template. We engineered fusions of CRISPR/Cas9 and a cytidine deaminase enzyme that retain the ability to be programmed with a guide RNA, do not induce dsDNA breaks, and mediate the direct conversion of cytidine to uridine, thereby effecting a C→T (or G→A) substitution. The resulting "base editors" convert cytidines within a window of approximately five nucleotides (nt), and can efficiently correct a variety of point mutations relevant to human disease. In four transformed human and murine cell lines, second- and third-generation base editors that fuse uracil glycosylase inhibitor (UGI), and that use a Cas9 nickase targeting the non-edited strand, manipulate the cellular DNA repair response to favor desired base-editing outcomes, resulting in permanent correction of ~15-75% of total cellular DNA with minimal (typically ≤ 1%) indel formation. Base editing expands the scope and efficiency of genome editing of point mutations.

We envisioned that direct conversion of one DNA base to another at a programmable target locus without requiring DSBs could increase the efficiency of gene correction relative to HDR without introducing an excess of random indels. Catalytically-dead Cas9 (dCas9), which contains Asp10Ala and His840Ala mutations that inactivate its nuclease activity, retains its ability to bind DNA in a guide RNA-programmed manner but does not cleave the DNA backbone.[7] In principle, conjugation of dCas9 with an enzymatic or chemical catalyst that mediates the direct conversion of one base to another could enable RNA-programmed DNA base editing.

The deamination of cytosine (C) is catalyzed by cytidine deaminases[8] and results in uracil (U), which has the base-pairing properties of thymine (T). Most known cytidine deaminases operate on RNA, and the few examples that are known to accept DNA require single-stranded (ss) DNA.[9]



Recent studies on the dCas9-target DNA complex reveal that at least nine nt of the displaced DNA strand are unpaired upon formation of the Cas9:guide RNA:DNA "R-loop" complex.[10] Indeed, in the structure of the Cas9 R-loop complex the first 11 nt of the protospacer on the displaced DNA strand are disordered, suggesting that their movement is not highly restricted.[11] It has also been speculated that Cas9 nickase-induced mutations at cytosines in the non-template strand might arise from their accessibility by cellular cytosine deaminase enzymes.[12] We reasoned that a subset of this stretch of ssDNA in the R-loop might serve as an efficient substrate for a dCas9-tethered cytidine deaminase to effect direct, programmable conversion of C to U in DNA (Fig. 1a).


dCas9
sgRNA
3'
5'
cytidine
deaminase
genomic DNA
binding
and opening
G
C
deamination
of target C
U
DNA replication
or repair
A
T
base-edited DNA
protospacer
PAM
genomic DNA

Base excision repair

In base excision repair, damaged heterocyclic DNA bases are detected and excised by DNA glycosylase enzymes. DNA glycosylases detect damaged bases in double-stranded DNA and excise them by hydrolysing the *N*-glycosidic bond. Different glycosylases target different types of DNA base damage. Figure 16 shows the mechanism of hydrolysis of the *N*-glycosidic bond of uridine nucleotides (uracil arises in DNA through deamination of cytosine) catalyzed by uracil DNA glycosylase (UDG). DNA glycosylases including UDG employ base flipping to enable them to reach the target base within the double helix.

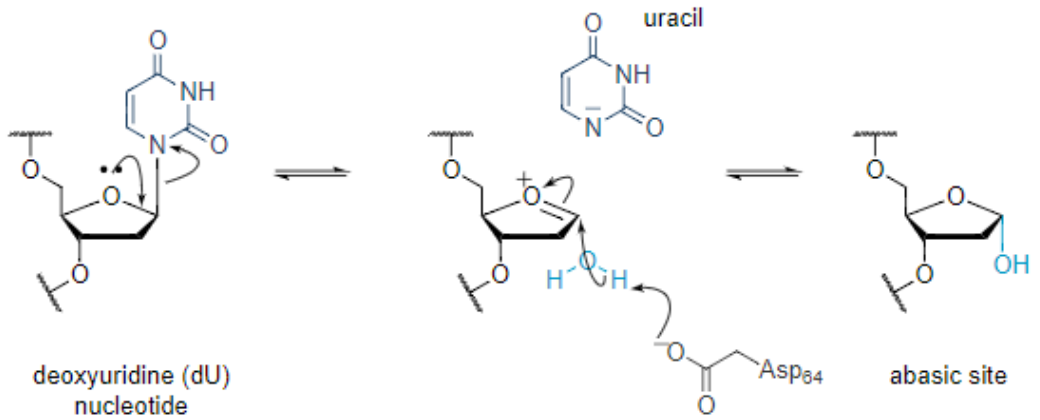
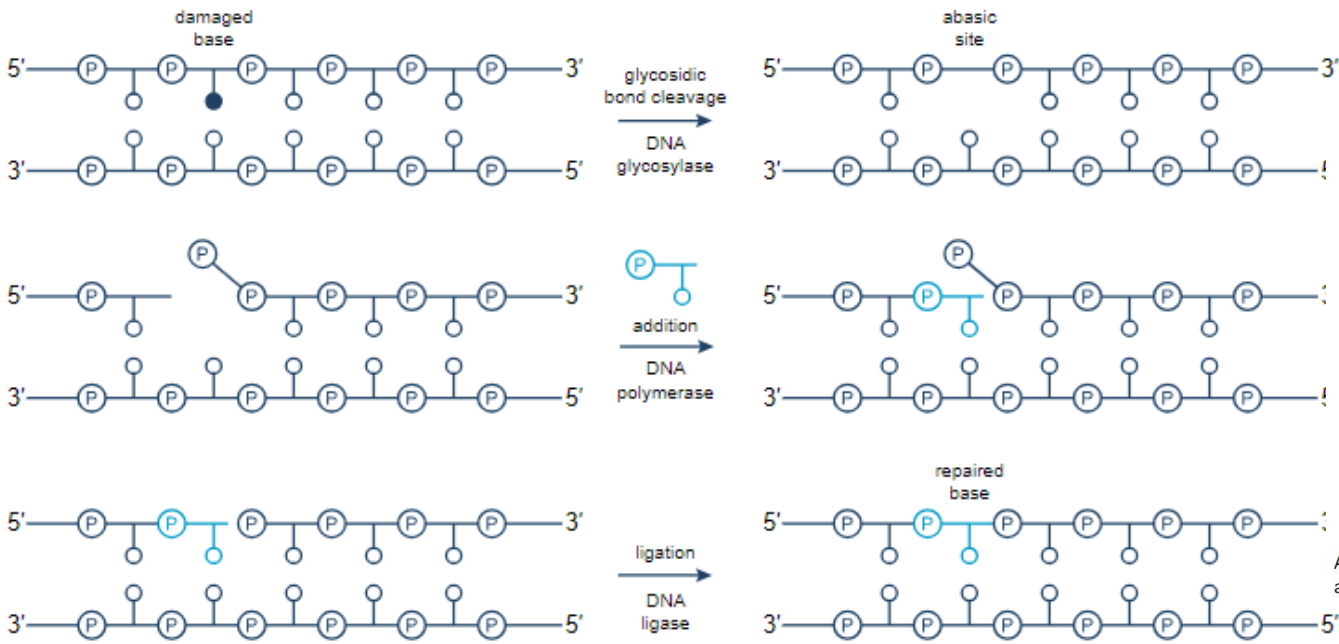
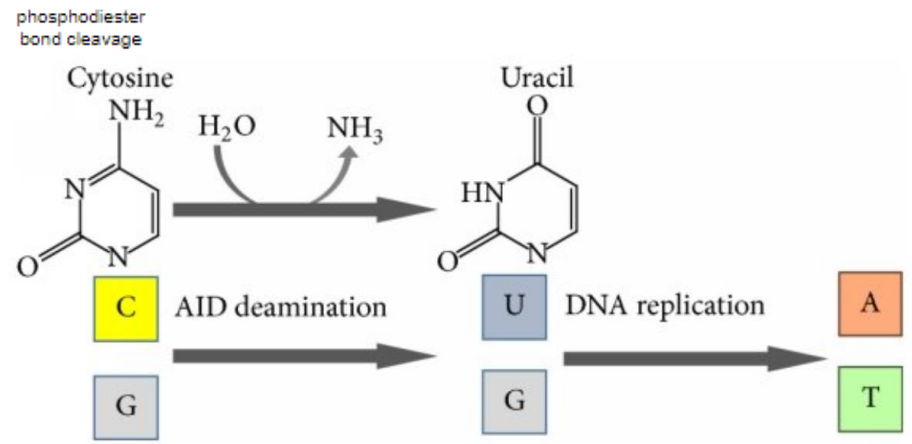


Figure 16 **UDG *N*-glycosidic bond hydrolysis** Mechanism of hydrolysis of the *N*-glycosidic bond of a uridine nucleotide, catalyzed by uracil DNA glycosylase (UDG), which initiates base excision repair of uridine nucleotides in DNA. The oxacarbenium ion intemediate is stabilized by electrostatic interactions with the enzyme (not shown).

Following excision of the damaged base, an abasic site (or AP site) is left in one strand of the double-stranded DNA. The abasic site is repaired in a series of steps by endonucleases, polymerases and ligases, using the complementary DNA strand as a template. Figure 17 shows an example of a base excision repair pathway.



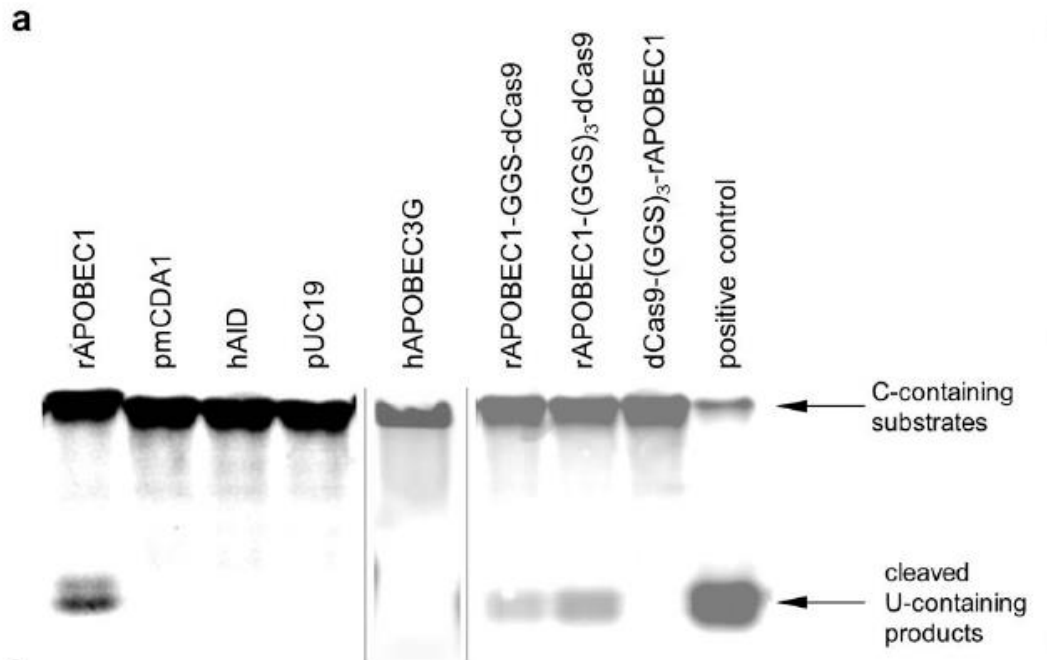


Activation-induced cytidine deaminase (AID) as the DNA/RNA editing enzyme. AID is induced by H. pylori infection and promotes the accumulation of various mutations.

Figure 17 **DNA base excision repair** The damaged base is recognized and excised by a DNA glycosylase, leaving an abasic (AP) site. An AP endonuclease cleaves the phosphodiester bond at the 5'-end of the damaged base. DNA polymerases add a nucleotide monophosphate to replace the damaged base, which is then removed by the lyase activity of the polymerase. Finally, a DNA ligase closes the gap to complete DNA repair. The mechanism of base excision repair varies between organisms, and several pathways can exist for a single organism.

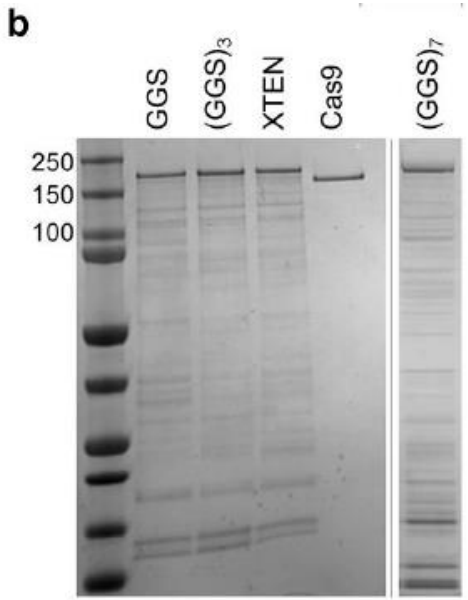
Four different cytidine deaminase enzymes (hAID, hAPOBEC3G, rAPOBEC1, and pmCDA1) were evaluated for ssDNA deamination. Of the four enzymes, rAPOBEC1 showed the highest deaminase activity under the conditions tested (Extended Data Fig. 1a). Fusing rAPOBEC1 to the N-terminus, but not the C-terminus, of dCas9 preserves deaminase activity (Extended Data Fig. 1a). We expressed and purified four rAPOBEC1-dCas9 fusions with linkers of different length and composition (Extended Data Fig. 1b), and evaluated each fusion for single guide RNA (sgRNA)-programmed dsDNA deamination *in vitro* (Fig. 1b and Extended Data Fig. 1c-f).

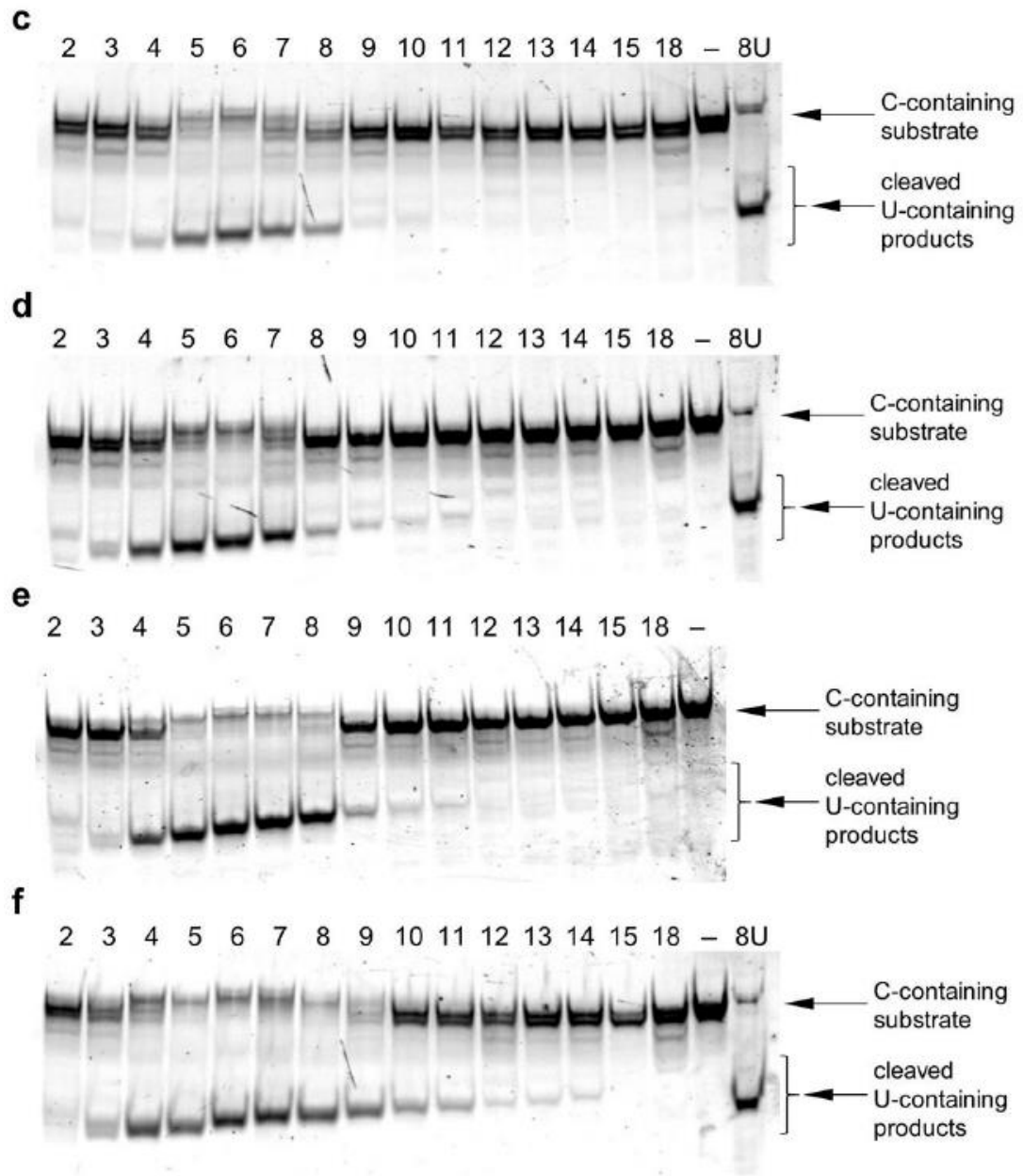
Extended Data



Extended Data Figure 1. Effects of deaminase, linker length, and linker composition on base editing

a, Gel-based deaminase assay showing activity of rAPOBEC1, pmCDA1, hAID, hAPOBEC3G, rAPOBEC1-GGS-dCas9, rAPOBEC1-$(GGS)_3$-dCas9, and dCas9-$(GGS)_3$-rAPOBEC1 on ssDNA. Enzymes were expressed in a mammalian cell lysate-derived *in vitro* transcription-translation system and incubated with 1.8 μM dye-conjugated ssDNA and USER enzyme (uracil DNA glycosylase and endonuclease VIII) at 37 °C for 2 h. The resulting DNA was resolved on a denaturing polyacrylamide gel and imaged. The positive control is a sequence with a U synthetically incorporated at the same position as the target C.





b, Coomassie-stained denaturing PAGE of the expressed and purified proteins used in (c), (d), (e), and (f). **c**-**f**, Gel-based deaminase assay showing the deamination window of base editors with deaminase–Cas9 linkers of GGS (c), $(GGS)_3$ (d), XTEN (e), or $(GGS)_7$ (f). Following incubation of 1.85 μM deaminase-dCas9 fusions complexed with sgRNA with 125 nM dsDNA substrates at 37 °C for 2 h, the dye-conjugated DNA was isolated and incubated with USER enzyme at 37 °C for 1 h to cleave the DNA backbone at the site of any Us. The resulting DNA was resolved on a denaturing polyacrylamide gel, and the dye-conjugated strand was imaged. Each lane is numbered according to the position of the target C within the protospacer, or with – if no target C is present. 8U is a positive control sequence with a U synthetically incorporated at position 8. For gel source data, see Supplementary Figure 1.

We observed efficient, sequence-specific, sgRNA-dependent C to U conversion *in vitro* (Fig. 1c). Conversion efficiency was greatest using rAPOBEC1-dCas9 linkers over nine amino acids in length. The number of positions susceptible to deamination (the "activity window") increases from approximately three to six nt as the linker length was extended from three to 21 amino acids (Extended Data Fig. 1c-f). The 16-residue XTEN linker[13] offered a promising balance between these two characteristics, with an efficient deamination window of approximately five nt, typically from positions 4 to 8 within the protospacer, counting the end distal to the protospacer-adjacent motif (PAM) as position 1. The rAPOBEC1-XTEN-dCas9 protein served as the first-generation base editor (BE1).

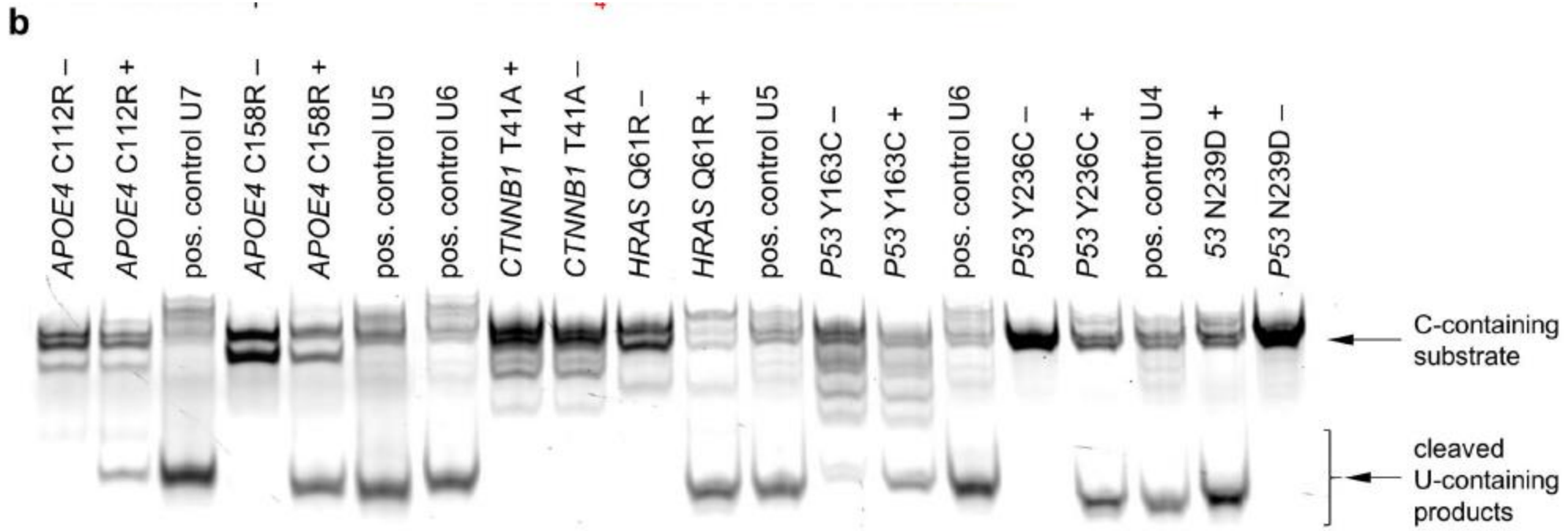
We assessed the ability of BE1 *in vitro* to correct seven T→C mutations relevant to human disease (Extended Data Fig. 2). BE1 yielded products consistent with efficient editing of the target C, or of at least one C within the activity window when multiple Cs were present, in six of these seven targets *in vitro*, with an average apparent editing efficiency of 44% (Extended Data Fig. 2).

a

APOE4 Cys112Arg:	5'-GGAGGACGTGC_{11}GCGGCCGCCTGG
APOE4 Cys158Arg:	5'-GAAGC_5GCCTGGCAGTGTACCAGG
CTNNB1 Thr41Ala:	5'-CTGTGGC_7AGTGGCACCAGAATGG
HRAS Gln61Arg:	5'-CCTCCC_6GGCCGGCGGTATCCAGG
P53 Tyr163Cys:	5'-GCTTGC_6AGATGGCCATGGCGCGG
P53 Tyr236Cys:	5'-ACACATGC_8AGTTGTAGTGGATGG
P53 Asn239Asp:	5'-TGTC_4ACACATGTAGTTGTAGTGG

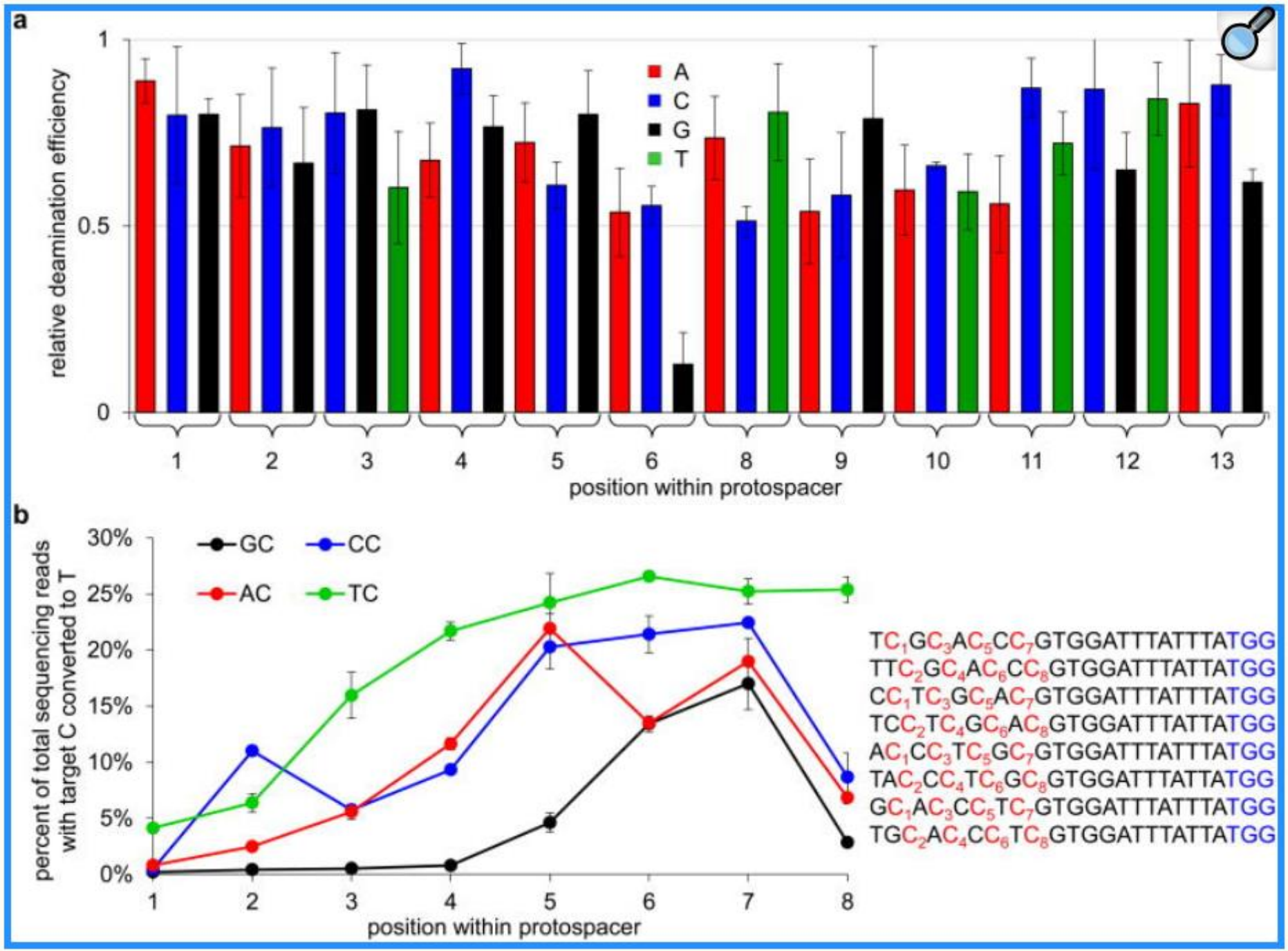
Extended Data Figure 2. BE1 is capable of correcting disease-relevant mutations *in vitro*

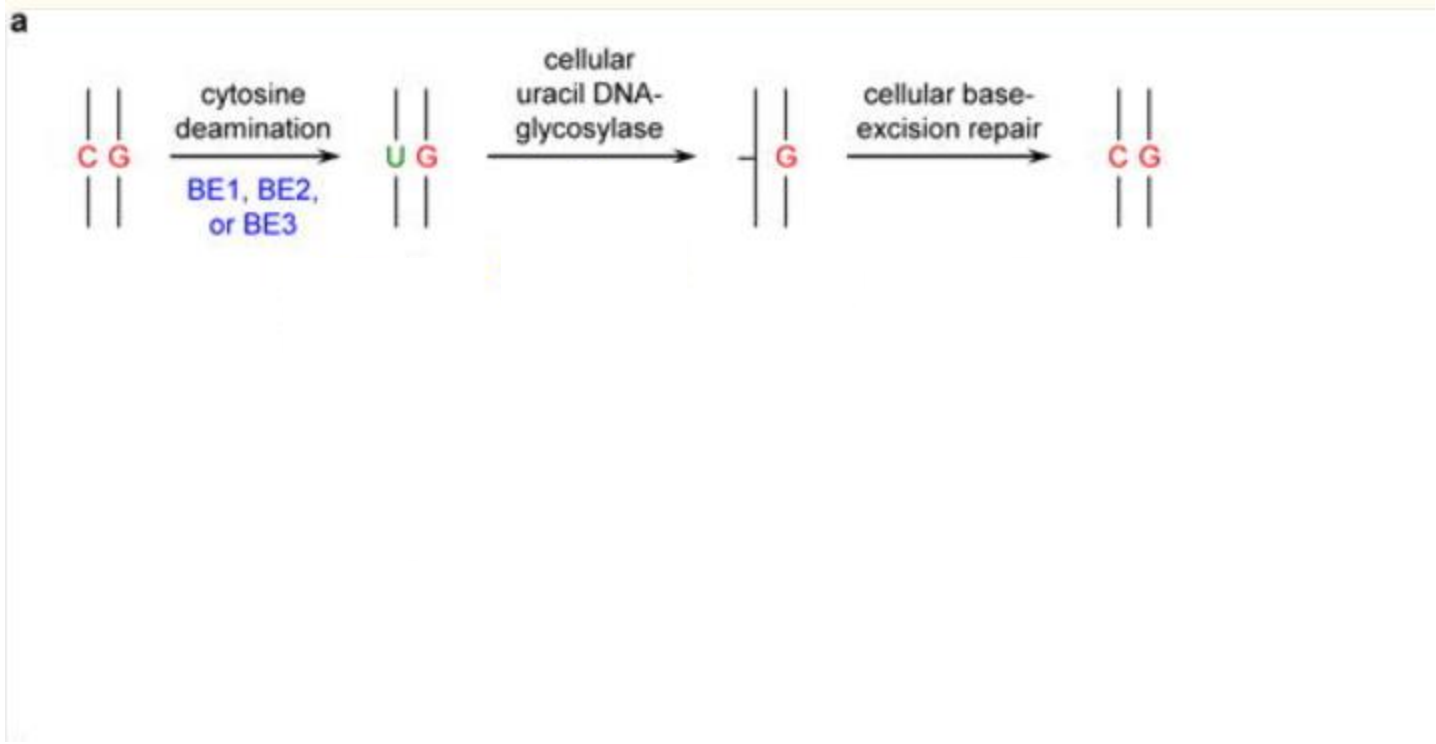
a, Protospacer and PAM sequences (red) of seven disease-relevant mutations. The disease-associated target C in each case is indicated with a subscripted number reflecting its position within the protospacer. For all mutations except both *APOE4* SNPs, the target C resides in the template (non-coding) strand.



b, Deaminase assay showing each dsDNA 80-mer oligonucleotide before (–) and after (+) incubation with BE1, DNA isolation, and incubation with USER enzymes to cleave DNA at positions containing U. Positive control lanes from incubation of synthetic oligonucleotides containing U at various positions within the protospacer with USER enzymes are shown with the corresponding number indicating the position of the U. Editing efficiencies were quantitated by dividing the intensity of the cleaved product band by that of the entire lane for each sample.

we assessed [Fig. 2b](). BE1 activity follo t C is at or near positio essive, and





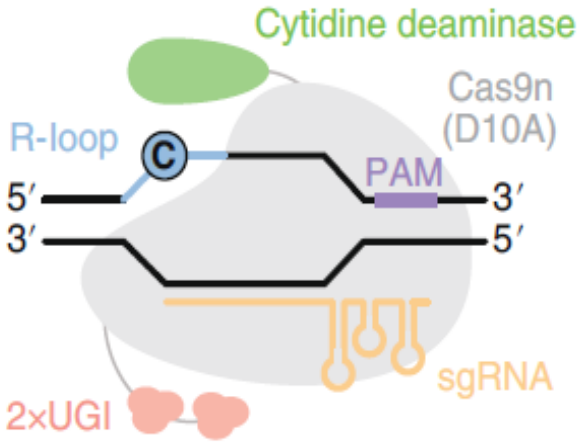
a, Possible base editing outcomes in mammalian cells. Initial editing results in a U:G mismatch. Recognition and excision of the U by uracil DNA glycosylase (UDG) initiates base excision repair (BER), which leads to reversion to the C:G starting state. BER is impeded by BE2 and BE3, which inhibit UDG. The U:G mismatch is also processed by mismatch repair (MMR), which preferentially repairs the nicked strand of a mismatch. BE3 nicks the non-edited strand containing the G, favoring resolution of the U:G mismatch to the desired U:A or T:A outcome.

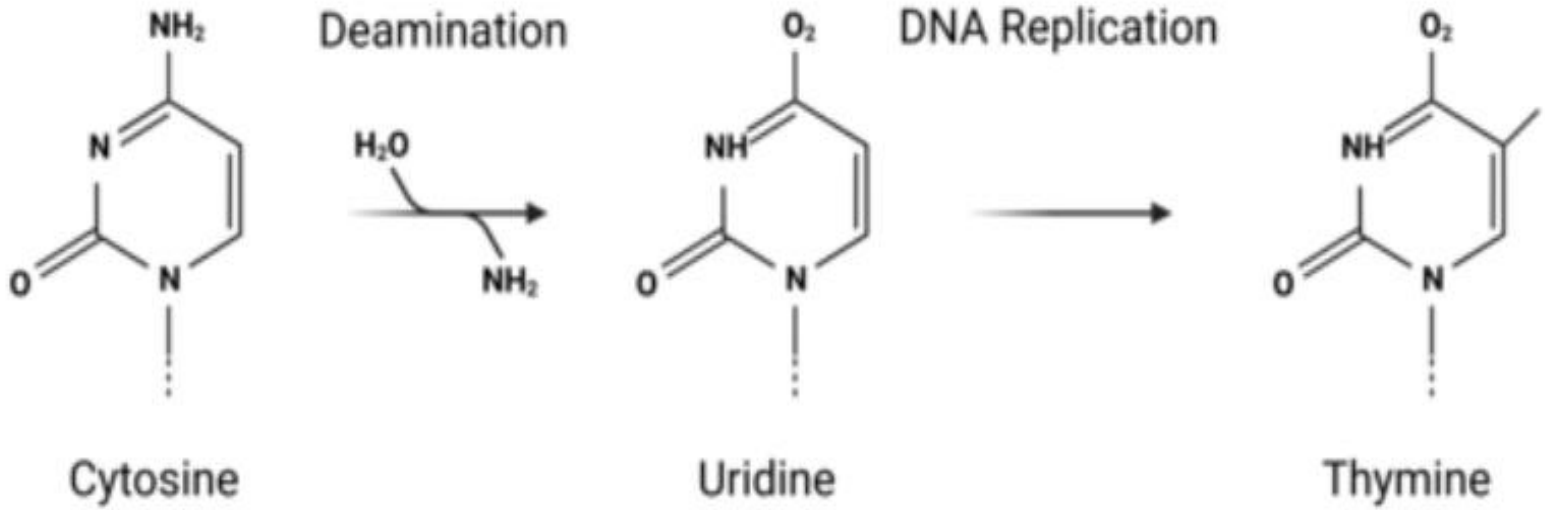
We hypothesized that the cellular DNA repair response to U:G heteroduplex DNA was responsible for the large decrease in base editing efficiency in cells (Fig. 3a). Uracil DNA glycosylase (UDG) catalyzes removal of U from DNA in cells and initiates base-excision repair (BER), with reversion of the U:G pair to a C:G pair as the most common outcome (Fig. 3a).[16] Uracil DNA glycosylase inhibitor (UGI), an 83-residue protein from *B. subtilis* bacteriophage PBS1, potently blocks human UDG activity (IC_{50} = 12 pM).[17] In an effort to subvert BER at the site of base editing, we fused UGI to the C-terminus of BE1 to create a second-generation base editor (BE2, APOBEC–XTEN–dCas9–UGI) and repeated editing assays on all six genomic loci. Editing efficiencies in human cells were on average 3-fold higher with BE2 than BE1, resulting in gene conversion efficiencies of up to 20% of total DNA sequenced (Fig. 3b).

Converting and protecting the substrate strand of a C:G base pair (bp) results in a maximum base editing yield of 50%. To augment base editing efficiency beyond this limit, we sought to further manipulate cellular DNA repair to induce correction of the non-edited strand containing the G. Eukaryotic mismatch repair (MMR) uses nicks present in newly synthesized DNA to direct removal and resynthesis of the newly synthesized strand (Fig. 3a).[19,20] We reasoned that nicking the DNA strand containing the unedited G would simulate newly synthesized DNA, inducing MMR to preferentially resolve the U:G mismatch into desired U:A and T:A products (Fig. 3a). We therefore restored the catalytic His residue at position 840 in the Cas9 HNH domain of BE2,[7] resulting in the third-generation base editor (BE3, APOBEC–XTEN–dCas9(A840H)–UGI) that nicks the non-edited strand containing a G opposite the edited U. BE3 retains the Asp10Ala mutation in Cas9 that prevents dsDNA cleavage, and also retains UGI to suppress BER.

What is base editing and what is it used for?

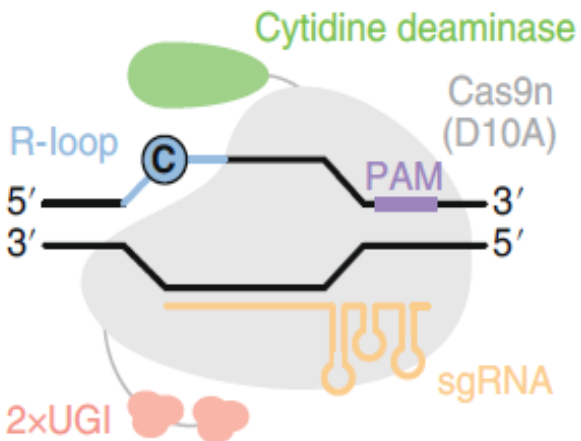
Base editing is a gene-editing technology described initially by the laboratories of Drs. David Liu and Akihiko Kondo in 2016 (1,2). Base editing combines the powerful DNA-scanning and sequence-identification capabilities of the CRISPR-Cas9 system with a deaminase enzyme, which introduces single nucleotide polymorphisms (SNPs) by chemically altering the target DNA sequence without the intentional generation of a DNA double-strand break (DSB). This chemical modification, known as deamination, consists of the removal of an amino group from a nucleotide, which after DNA repair or replication results in the installation of a new base

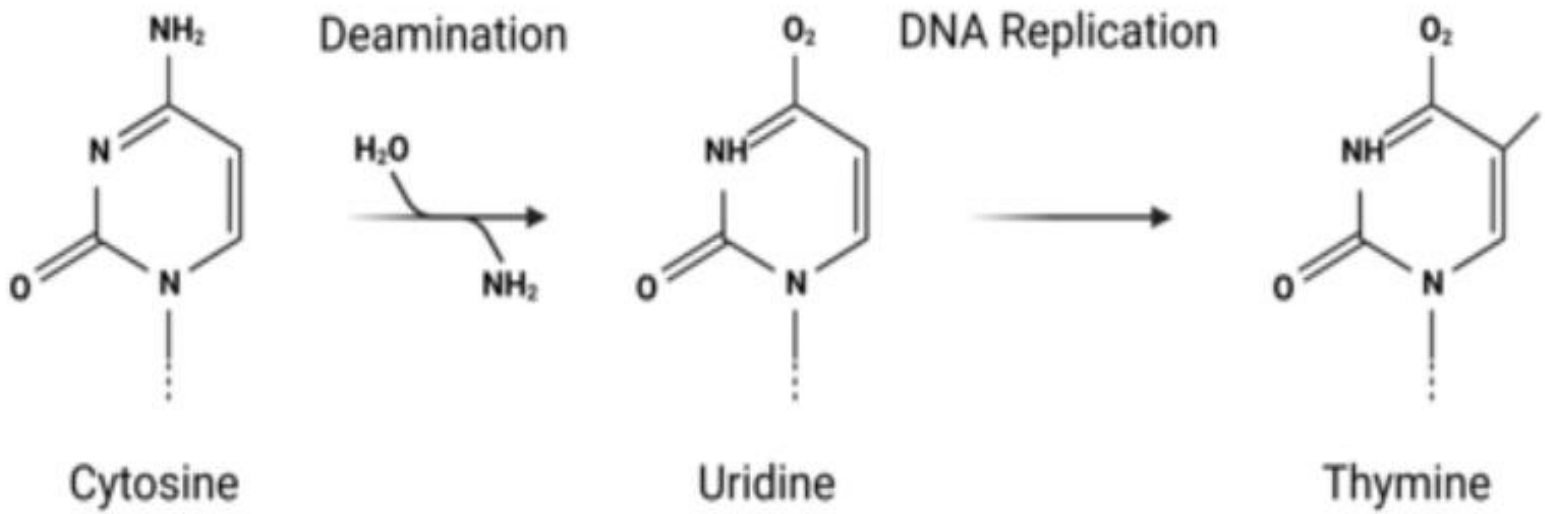


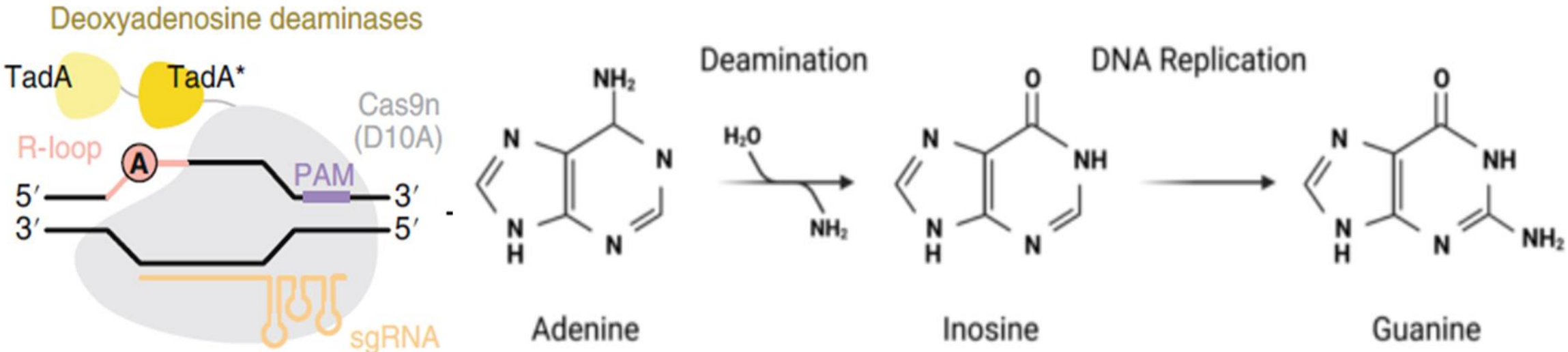


What is base editing and what is it used for?

The first base editors developed in 2016, called cytosine base editors (CBEs), can substitute a cytosine base for a thymine (C→T), while newer base editors called adenine base editors (ABEs) are able to introduce adenine-to-guanine (A→G) substitutions (3). More recently, base editors capable of cytosine-to-guanine (CGBEs) or simultaneous adenine and cytosine (ACBEs) substitutions have also been developed but their application is still limited to a handful of studies (4,5).







Base editing is an attractive way to edit DNA, because it greatly reduces some of the major risks associated with conventional, DSB-mediated gene editing: the heterogeneity of insertions and deletions (INDELs) at on-target sites, the increased potential for DSB-related chromosomal aberrations, and cell toxicity.

What are the components of a base editor?

Base editors are generated by coupling two separate proteins capable of very specific functions, as illustrated in Figure 2. The components of a typical base editor include:

- **Cas9 nickase (nCas9):** This is a version of Cas9 that has been modified through mutations in one of the two main amino acid residues responsible for the DNA cleavage activity of Cas9. Thus, nCas9 is still able to pair with a gRNA and find the DNA sequence complementary to the gRNA spacer, but is only able to nick one strand of the DNA.
- **Nucleoside deaminase:** This is an enzyme capable of removing an amino group from a specific type of nucleoside. The deaminase fused to the nickase – be it a cytosine deaminase (e.g. APOBEC) or an adenosine deaminase (e.g. engineered TadA) – will dictate whether the base editor is a CBE or an ABE. CBEs contain cytosine deaminases, while ABEs typically contain adenine deaminases (though several groups have recently engineered CBEs from the conventional TadA adenosine deaminase domain (6-8)).

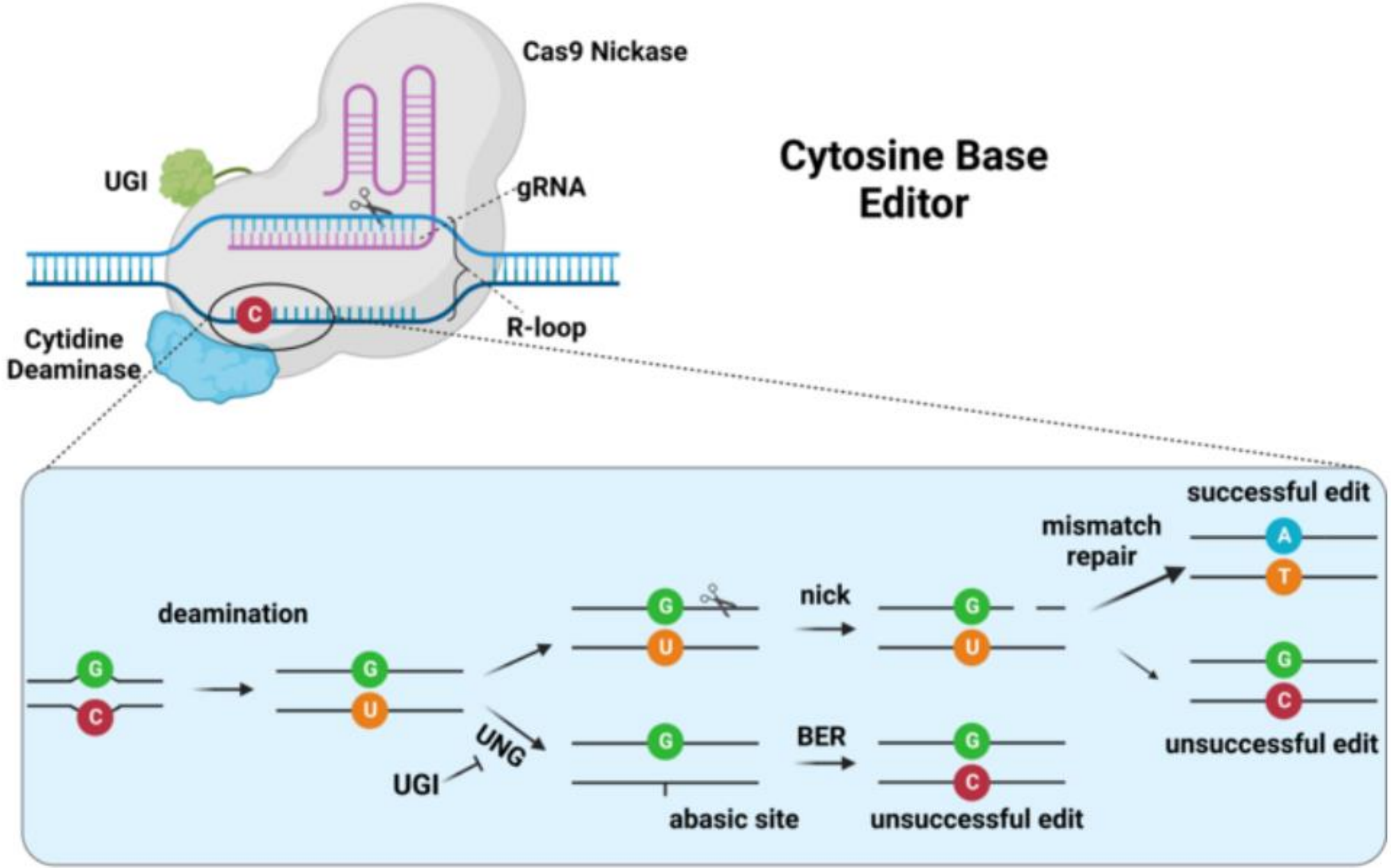
The Cas9-deaminase fusion protein complex is targeted to a specific DNA locus by a **guide RNA (gRNA)**. Once the base editor has bound to its target sequence, the deaminase can modify bases within the exposed non-target strand (NTS) of the target site R-loop. The area of the target locus in which bases can be modified is called the **base editing window** (see Figure 2). Depending on the specific intended modification, other factors may be included in the base editor complex to increase efficiency or alter the outcome of the base substitution.

How does base editing work?

Once the base-editing complex recognises the target DNA sequence complementary to the gRNA spacer, Cas9 mediates the denaturing of the double-stranded DNA (dsDNA) to form a structure known as an R-loop, which exposes a short stretch of single-stranded DNA (ssDNA) in the non-complementary strand to the activity of the deaminase (see Figure 3; also called the non-target strand). This allows the deaminase to chemically modify bases inside the base editing window, which is normally 5-10bp long depending on the deaminase, and somewhat distal from the protospacer adjacent motif (PAM) within the target site.

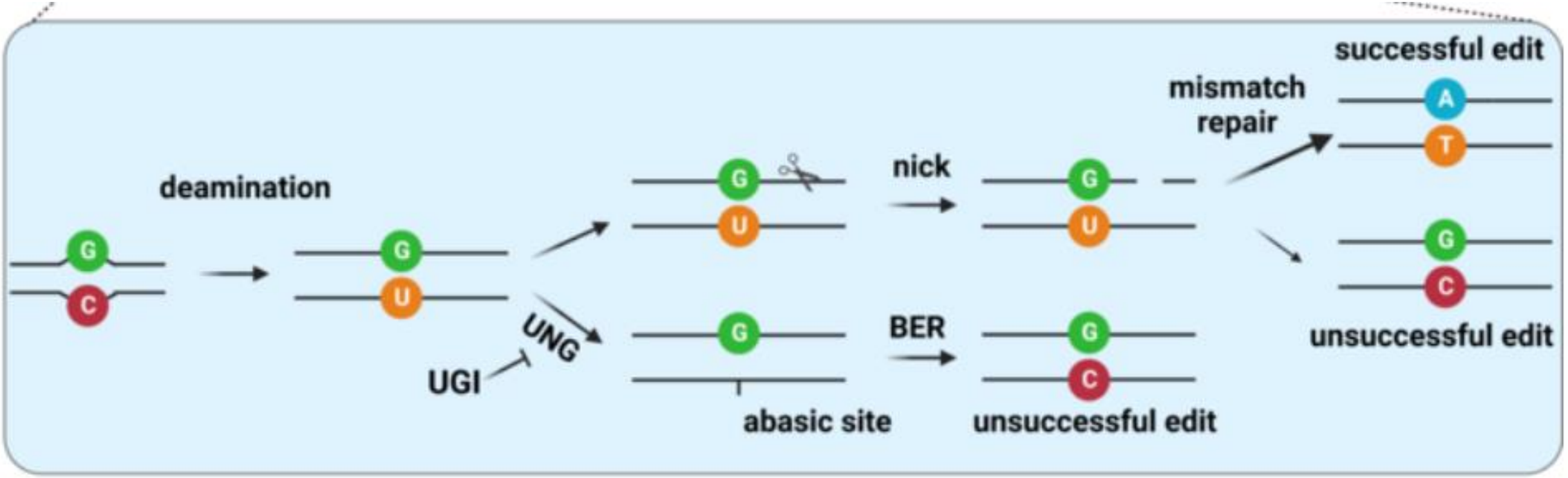
The first generation of base editors encountered severe efficiency issues that had to be solved. For example, **CBEs** mediate the deamination of cytosines, generating a uracil (U) base, which can then be read by the cell's DNA polymerases as thymine. However, human cells are very efficient at DNA repair, and especially at dealing with deaminated bases via the base-excision repair (BER) pathway. Thus, uracil can be recognised and eliminated by the enzyme uracil DNA N-glycosylase (UNG) during the initiation of BER. This drastically limits the efficiency of C-to-T edits and can lead to reduced C-to-T purity (and increased C-to-A or C-to-G edits). To overcome this issue, second-generation CBEs included a third component, namely a uracil glycosylase inhibitor (UGI) fused together with Cas9 and the cytosine deaminase. UGI inhibits the action of UNG on the edited DNA, thus conserving the uracil base and achieving 3-fold higher editing efficiency and purity (Ref. 1, and see Figure 3).

How does base editing work?



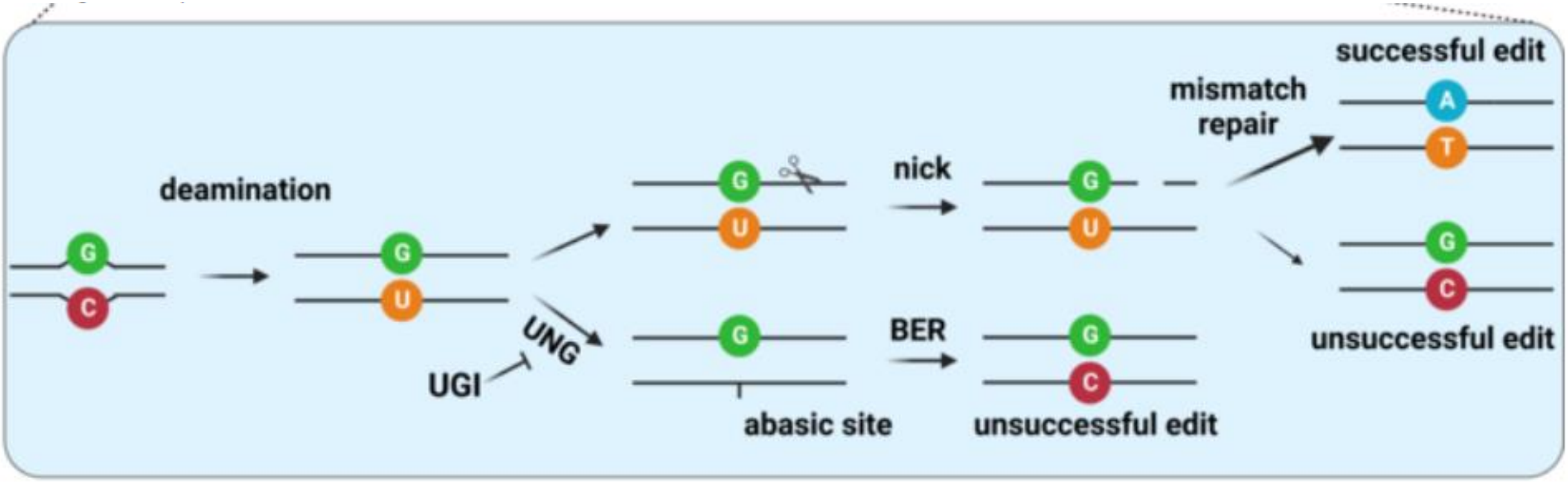
How does base editing work?

In addition to the above, the DNA mismatch repair mechanisms of the cell are tasked with detecting and resolving any incorrect nucleotide pairings into correct pairings (A-T and C-G). Since the deamination of a cytosine, which will be paired with a guanine, will lead to an incorrect T-G pairing, this will be repaired by the cell, either by substituting the U for a C (C-G pairing) or by substituting the G for an A (T-A pairing).



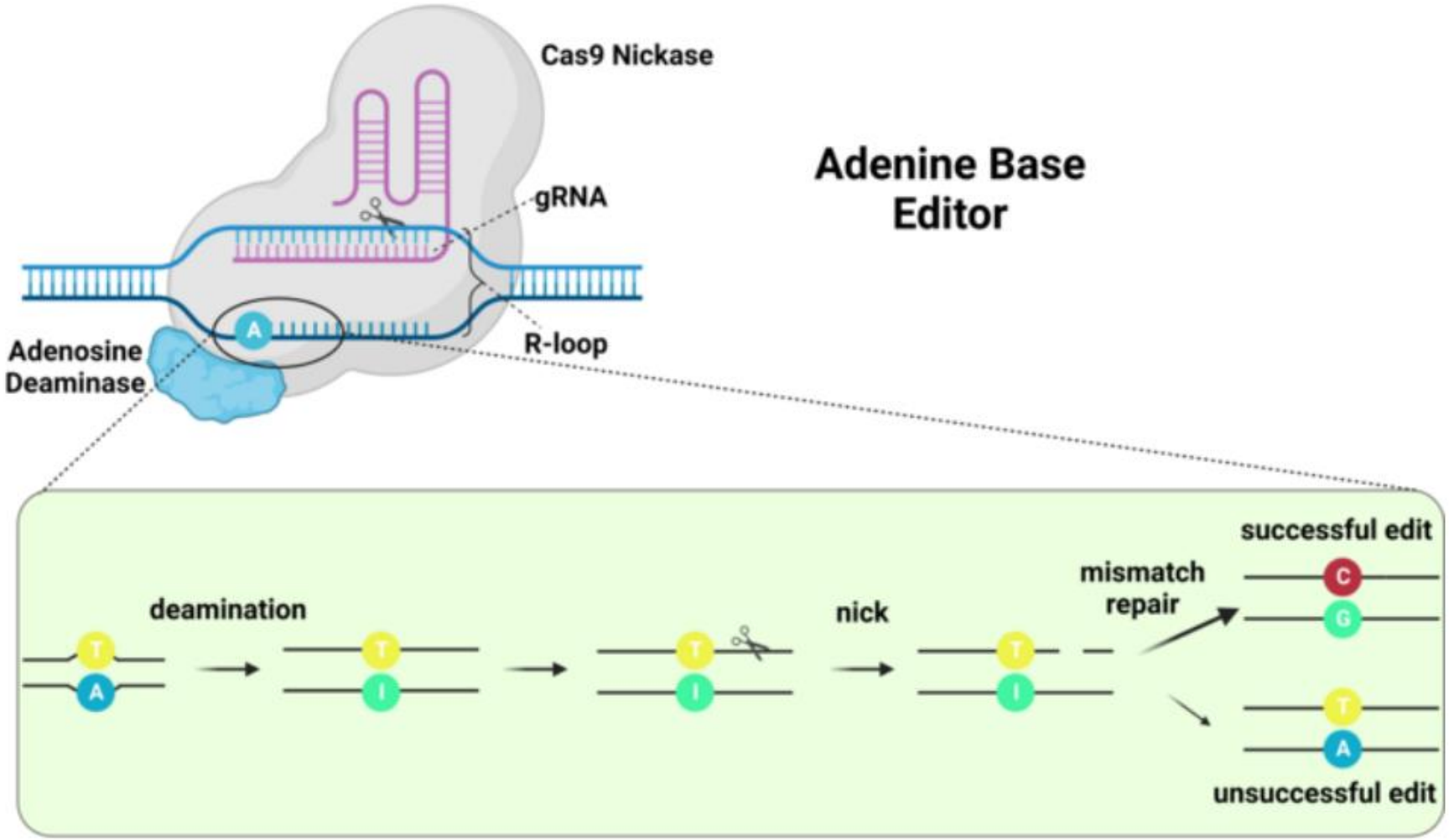
How does base editing work?

Considering that the cell does not know which of the two bases is correct, the probabilities of each repair outcome are equal. To increase editing efficiency, the catalytically-dead version of Cas9 (dCas9) that was used in the first-generation CBEs was substituted with the Cas9 nickase (nCas9) version that we find nowadays in nearly all base editors, which is only able to nick the complementary strand of the DNA. This cut in the complementary (=unedited) strand, directs the mismatch repair machinery to utilise the base-edited strand as a template for repair, thus biasing the repair towards the introduced edit, and further improving efficiency



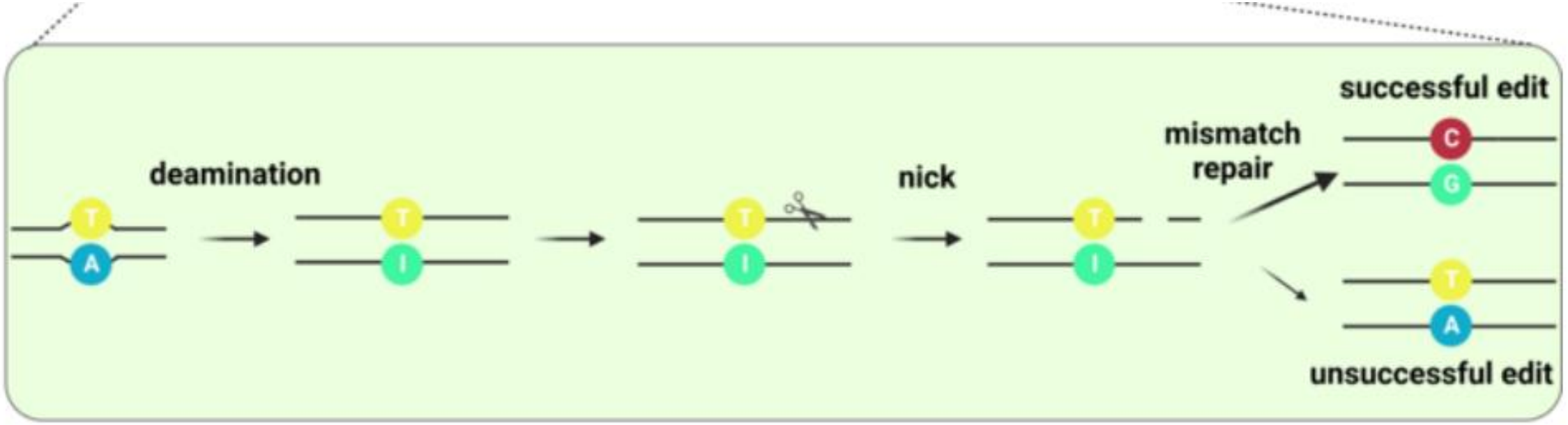
How does base editing work?

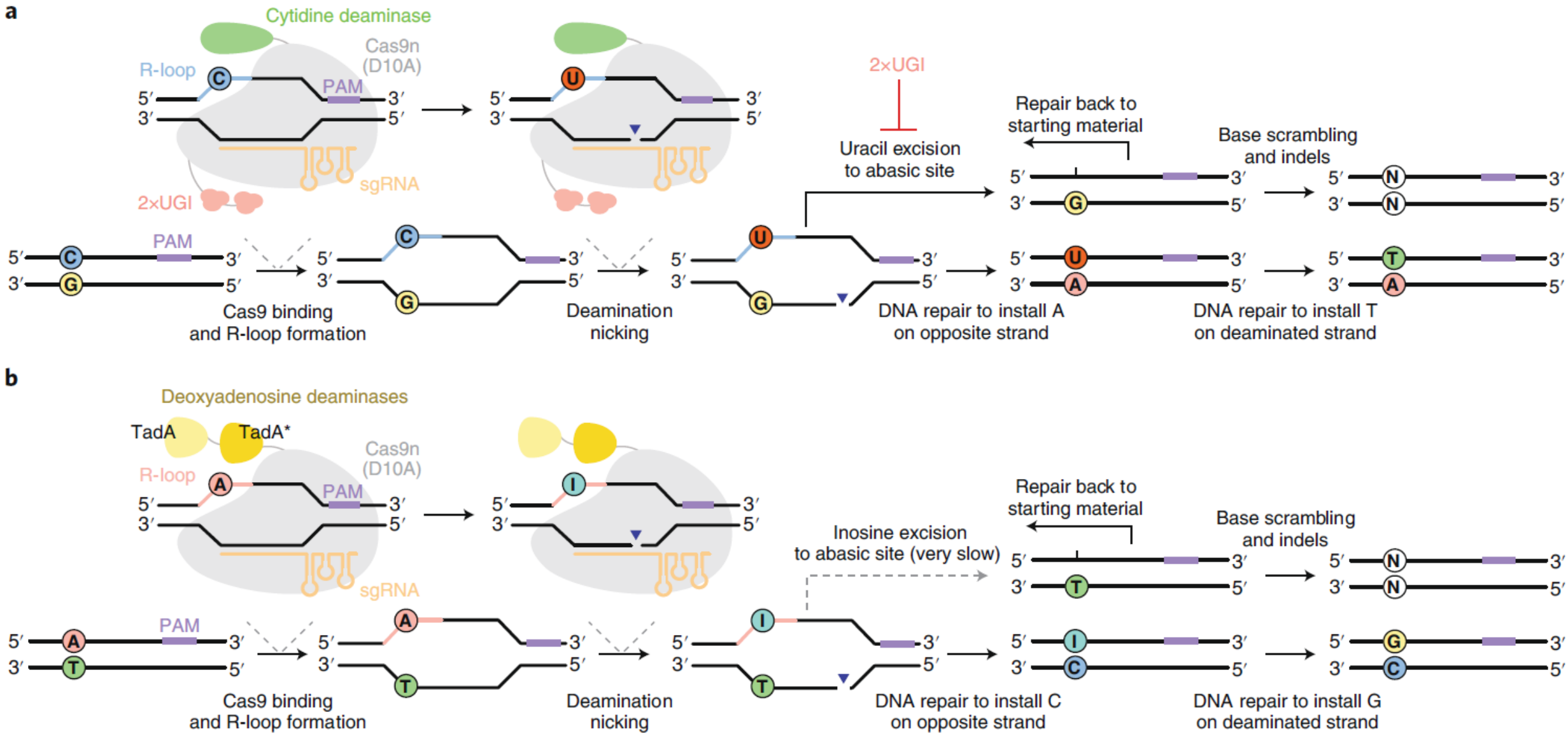
ABEs work in a very similar way. In this case, the deamination of adenine generates an inosine residue, which is interpreted by the DNA polymerases as a guanosine. Thus, an A-to-G edit is achieved (Ref. 3, and see Figure 4). However, no UGI or equivalent inhibitor is necessary since inosine residues are not excised like uracil residues.



How does base editing work?

One step that was very important for the generation of ABEs was making an adenine deaminase that could edit the DNA, since naturally occurring adenine deaminases only edit RNA. Directed mutagenesis efforts were successful in generating a version of the *E.coli* tRNA adenosine deaminase (TadA) that can edit the DNA. Interestingly, editing efficiency was higher when the ABEs included one mutated TadA (TadA*) and one wildtype enzyme (3). ABEs are very important because G-C to A-T mutations represent the most frequently reported pathogenic point mutations, thus this type of base editing has the greatest potential for clinical applications. More recent ABEs have been developed that have improved editing efficiencies (9,10).




a
Cytidine deaminase
Cas9n
(D10A)
R-loop
PAM
sgRNA
2×UGI
Cas9 binding
and R-loop formation
Deamination
nicking
2×UGI
Uracil excision
to abasic site
Repair back to
starting material
Base scrambling
and indels
DNA repair to install A
on opposite strand
DNA repair to install T
on deaminated strand
b
Deoxyadenosine deaminases
TadA
TadA*
Cas9n
(D10A)
R-loop
PAM
sgRNA
Cas9 binding
and R-loop formation
Deamination
nicking
Inosine excision
to abasic site (very slow)
Repair back to
starting material
Base scrambling
and indels
DNA repair to install C
on opposite strand
DNA repair to install G
on deaminated strand

How has base editing been used so far?

Several fascinating therapeutic approaches using base editing have been developed all over the world in the last few years. For example, in August 2022, Verve Therapeutics initiated a clinical trial using an intravenously delivered lipid nanoparticle that carries an ABE mRNA and a gRNA programmed to disrupt and silence the *PCSK9* gene in the liver of patients with familial hypercholesterolemia (11).

More recently, two independent research groups have shown the potential of base editing to treat spinal muscular atrophy in patient-derived cells and mice by introducing an A-to-G edit in the *SMN2* gene to increase its expression and compensate for the lack of expression in the mutated paralogous *SMN1* gene (12,13).

Another interesting use of base editing has been to simultaneously disrupt the CCR5 and CCRX4 receptors of CD4+ T cells. This has been shown to be an effective method of rendering those cells immune to HIV infection and has the potential to become an effective way to prevent HIV infection/reinfection (14).

Although major advances have been made within base-editing technology so far, we have likely only scratched the surface of its potential in research and therapeutics.


Welcome to Robin Ward
Robin

What are the limitations of base editing?

Some of the limitations are similar to those of conventional Cas9-based editing approaches regarding targeting, off-target editing and delivery. Thus, already existing Cas9 variants that were developed to solve those issues can be used to construct base editors able to overcome the same issues. When it comes to *in vivo* delivery, the increased size of the base editor complex compared to regular Cas9 complicates its delivery. To address this hurdle, the intein system, which is used to package other Cas9 products into adeno-associated viral vectors, has also been used for *in vivo* delivery of base editors. However, other, base editing-specific limitations have arisen:

- **Requirement for precise positioning of the base editor edit window:** To achieve efficient editing, base editors must be capable of recognising PAMs at an appropriate distance upstream of the intended edit (to position the edit window accordingly). Fortunately, various Cas9 PAM variants have been developed that utilise alternate PAMs, enabling more flexible base editing across the genome (15-17).
- **Unwanted edits in the editing window:** In some cases, when trying to edit a specific base, for example a cytosine, another cytosine inside the editing window can be deaminated as well, leading to unwanted effects. To avoid this, base editor variants with narrower editing windows have been developed by introducing mutations in the deaminase enzymes.

What are the limitations of base editing?

- **Indel generation:** Since new-generation base editors introduce single-strand nicks in the non-edited strand, in some cases the excision of an edited base by the BER can lead to a DSB. The repair of DSBs leads to the generation of INDELs. Strongly inhibiting the BER pathway is one approach to avoid DSB formation.
- **Off-target RNA editing:** An interesting and unexpected observation made by some scientists was that base editors could also produce unwanted edits to RNA, independently of Cas9 activity, which could lead to unpredictable effects. Efforts to solve this issue have focused on engineering the deaminase protein to reduce its activity on the RNA.

Thank You!