Title	遺伝子修復と遺伝子治療のための人工RNA編集システムの改良
Author(s)	李, 嘉睿
Citation	
Issue Date	2023-06
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/18709
Rights	
Description	Supervisor:芳坂 貴弘, 先端科学技術研究科, 博士

Doctoral Dissertation

Improvement of artificial RNA editing system for genetic restoration and gene therapy

Jiarui Li

Supervisor: Takahiro Hohsaka

Graduate School of Advanced Science and Technology

Japan Advanced Institute of Science and Technology

Materials Science

June 2023

Abstract

The enzymes in the adenosine deaminase acting on RNA (ADAR) family have a deaminase domain (DD) that converts adenosine (A) into inosine (I), which functions as guanosine (G) during translation. In diseases caused by point mutations in genes, an important method for correcting RNA sequences and ultimately fine-tuning protein function is the artificial site-directed RNA editing.

In this study, I attempted to extend the MS stem-loop RNA to bind DD and guide RNA for the purpose of the editing efficiency of the MS2-ADAR1 RNA editing system. But the replacement of 6 X MS2 stem-loop RNA with 12 X MS2 stem-loop RNA was not valid due to the distance between the antisense part and the stem-loop part. My colleagues have developed a guide RNA that inserts stem-loops on both side of the complementary sequence of the target RNA, so I tried to improve the editing efficiency base on the 1-1 stem-loop guide RNA system. I increased the number of stem-loop on both sides and changed the paired base of the target nucleotide. The results showed that when the number of stem-loop on both sides was the same, the system showed high editing efficiencies in all conditions. In case of the paired base, when the paired base was U, the editing efficiency of this system was higher than other bases. These improvements might be very useful for treating genetic diseases that result from the G to A point mutation.

Our lab also developed the MS2-APOBEC1 system for restoring the T to C point mutation. It was used the deaminase domain of APOBEC1 (apolipoprotein B mRNA editing catalytic polypeptide 1) linked to the MS2 coat protein to perform the C to T deamination. I replaced the APOBEC1 catalytic domain with full-length human APOBEC3A and APOBEC3G. However, I can't detect any fluorescent signal from the cells transfected with the original guide RNA and APOBEC3A or APOBEC3G. I referred the natural substrate of APOBEC3A and APOBEC3G, so I designed the loop guide RNA for inducing loop structure on the target RNAs. I designed six types of guide RNAs to induce different lengths of loops for comparing the editing efficiencies. In all guide RNA conditions, I found that the 14nt loop guide RNA transfected with MS2-APOBEC3A or MS2-APOBEC3G could induce higher RNA editing efficiencies. I also used the D317W mutation of APOBEC3G transfected with the loop guide RNA. Even the D317W mutated APOBEC3G showed sequence preference to 5'-TC, the editing efficiencies were increased slightly. However, I couldn't detect any fluorescent signals from the cells that were transfected by the loop guide RNA and MS2-APOBEC1 system. For the application of guide RNA, it needs further optimization for improving the editing efficiency of the MS2-APOBEC system.

The proper application of the developed artificial deaminase system for the treatment of patients who have G-to-A or T-to-C mutations could open a new era in the field of treatment of genetic diseases.

Keywords: artificial RNA editing, base deamination, ADAR1, APOBEC, MS2 RNA

Table of Contents

Chapter3 Improvement of MS2-ADAR system for site-directed RNA editing 69

Chapter1 General Introduction

1.1 Genetic engineering

Genetic engineering, also called as genetic modification or transgenic technology, is a biotechnology that involves direct manipulation of an organism's genome and alteration of the genetic material of cells [1]. This technology encompasses both intra-species and cross-species gene transfer for the purpose of creating improved or novel organisms [2]. There are several methods for inserting new genetic material into the host genome, including isolation and replication of target genetic material using molecular cloning method to generate DNA fragments, or synthesis of DNA followed by insertion into the host's genome. Conversely, genes can also be removed or "knocked out" using nucleases [3]. Additionally, gene targeting, was also utilized to remove exons, add genes, or create point mutations [4].

1.2 Genome editing

Genome editing, also referred to as gene editing, is a type of genetic engineering that involves manipulating the living genome through the insertion, deletion, modification, or replacement of DNA. [5].

Genome editing technology has been in use since the 1970s, but its unpredictability in DNA insertion into the host genome poses a risk of harming or altering other genes in the organism [6]. The fundamental principles of modern genome editing technology involve the activation of the cell's natural repair mechanism through specific DNA double-strand breaks (DSB), which include two pathways of non-homologous end joining (NHEJ) and

homologous directed repair (HDR) [7]. However, NHEJ may lead to random base insertions or deletions during DNA repair and reconnection, resulting in frameshift mutations and gene inactivation to achieve knockdown of the target gene. Site-directed gene knock-in is made possible by the NHEJ mechanism, which ligates an external donor gene sequence into the DSB site of the double-strand break [8-14].

Homologous directed repair (HDR) is a comparatively high-fidelity repair method that involves the integration of an exogenous target gene in the donor into the target site through the homologous recombination process in the presence of a recombinant donor with homologous arms, without random base insertion or deletion. If DSBs are produced on both sides of a gene, the original gene can be replaced while a homologous donor is present [15-21]. Zinc-finger nucleases (ZFNs) [22], RNA-guided engineered nucleases (RGENs) [23], transcription activator–like effector nucleases (TALENs) [24], and the famous CRISPR/Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats) played important roles in DNA editing [25-26]. Despite this, DNA editing methods DNA editing, which may make permanent changes to the genome, have their own disadvantages. Whereas RNA editing does not cause such risk outcomes.

In my doctoral thesis, I mainly focused on RNA editing.

1.3 RNA editing

RNA editing is a process that modifies the nucleotide sequence of RNA after transcription. This process can involve additions, deletions, modifications, or conversions of nucleotides within the coding region of the RNA molecule [27, 28]. Originally, the nomenclature of "RNA editing" referred to the description of the insertion of four uridines

into the mitochondrial transcript in *Trypanosoma brucei*. However, the term has since been broadened to encompass a variety of procedures that modify the nucleotide sequence of RNA [29]. RNA editing phenomena in mammals were first found in 1987 [30]. In recent years, it has been considered a powerful genetic engineering tool that can effectively modify the base sequence of RNA after DNA editing.

RNA editing technology has recived significant attention due to its potential for treating genetic diseases caused by single nucleotide mutations. Base conversion, particularly A-to-I and C-to-U, is the most widely used method in RNA editing (**Fig. 1**) [31, 32]. Besides, my colleagues discovered the conversion of U-to-C in *Arabidopsis thaliana* (**Fig. 2**) [33, 34]. Compared to DNA editing, RNA editing is a safer approach in theory. By modifying the coding sequence of a gene post-transcriptionally, RNA editing was allowed to produce different types of proteins from the same gene [35]. The mature mRNA contains only exons. Thus, the intron editing is not meaningful. Meanwhile, since the mRNA will degrade automatically after the translation, no mutations could transmit to offspring [36-38].
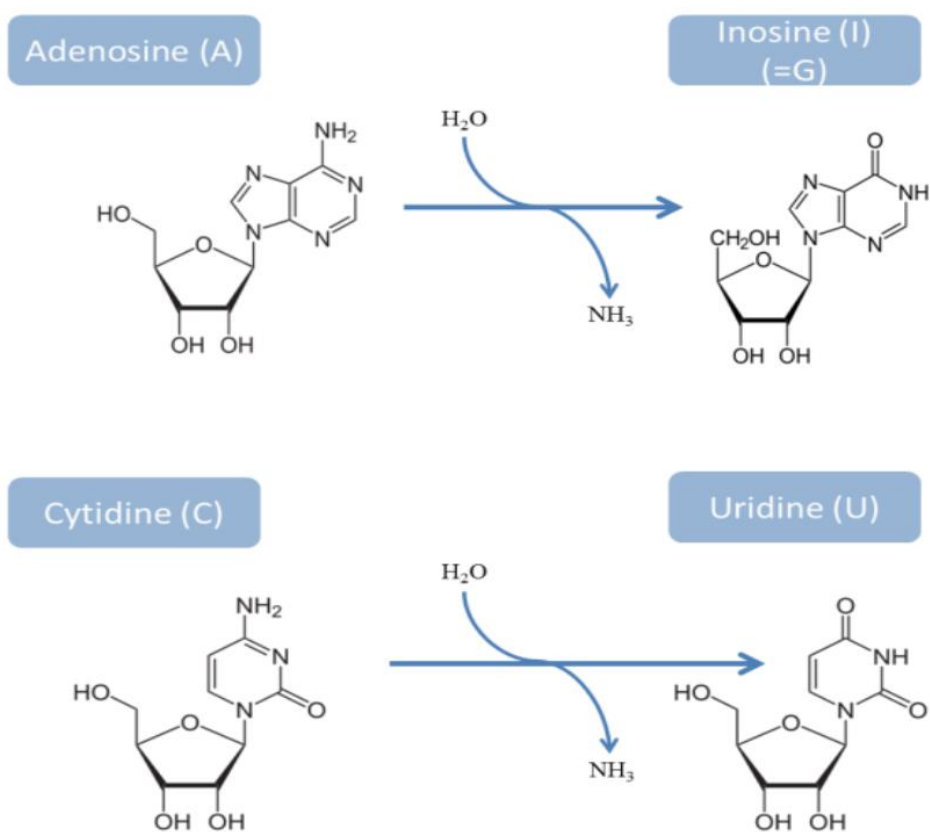
Fig 1: Chemical conversion of A-to-I and C-to-U. (Nucleotide base figures are reference from Wikipedia)
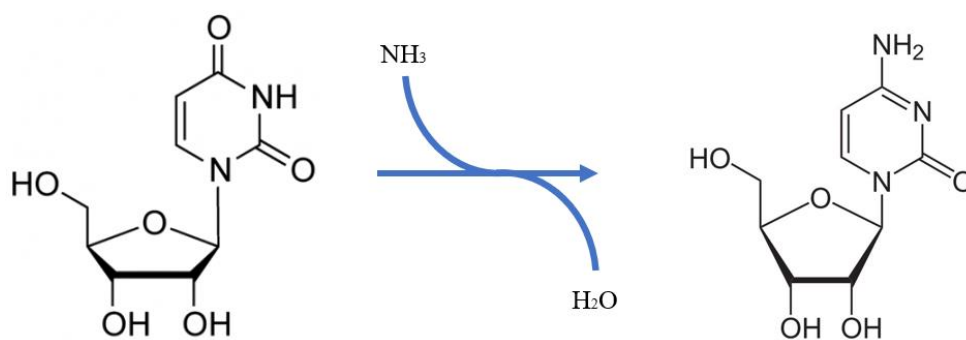
Fig 2. Chemical conversion of U-to-C (Nucleotide base figures are reference from Wikipedia)

1.4 Mechanism of C to U (T)editing

The cytidine deaminases that can accomplished by activation-induced cytidine deaminase (AID) on DNA level and apolipoprotein B mRNA editing catalytic polypeptide-like (APOBEC) family members on DNA or RNA level, respectively [39]. APOBEC1 is one of the primary C-to-U deaminase tools for mRNA editing. However, under hypoxic or overexpression conditions, APOBEC3A and APOBEC3G can also play important roles in RNA editing [40-42]. All these previous findings suggested that additional APOBECs could potentially function as RNA editing enzymes.

The most well described enzyme was the human APOBEC1, which was first identified as the enzyme take charge of the RNA editing of the APOB mRNA in the small intestine (**Fig. 3**). This editing event leads to the synthesis of the protein APOB-48, which is indispensable for the assembly and secretion of chylomicrons [30, 43, 44]. Human body requires both isoforms, which are generated from the same mRNA, for lipid metabolism. The conversion of the cytidine of the glutamine codon (CAA) to the uridine stop codon TAA at position 6666 results in the shorter variant. This conversion and the resulting creation of the smaller apoB48 (241 kDa) are also necessary for the body to absorb dietary lipids [37, 43]. The longer version (512 kDa), which is generated in the liver, is responsible for the transporting endogenously produced triglycerides and cholesterol in the bloodstream. In C-to-U RNA editing, the mooring sequence is crucial for site selection.
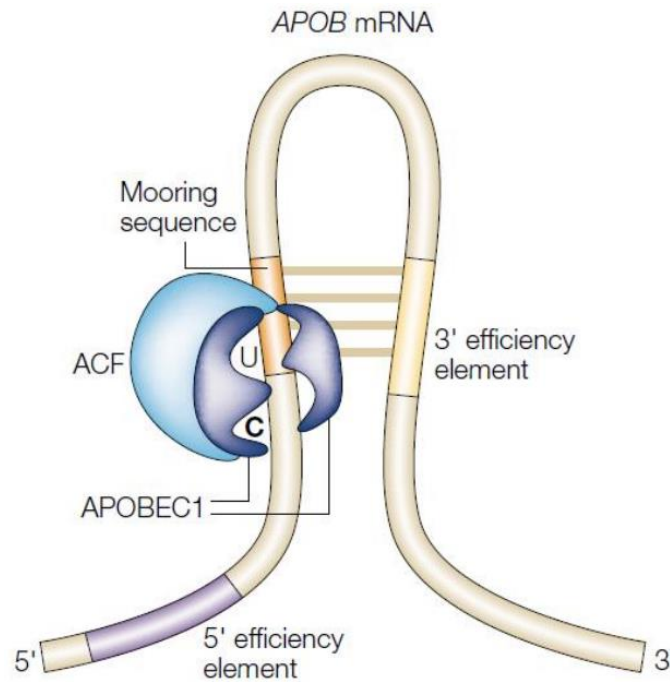
Fig 3: With the aid of the associated factor (ACF), APOBEC1 attaches close to the Mooring

region of APOB mRNA and deaminates C-to-U. [27].

1.5 Mechanism of A to I editing

A-to-I RNA editing events is the most common form in vertebrates [45, 46]. In this process, adenosine can be converted to inosine, whereas inosine serves as guanosine during the translation. RNA editing through adenosine (A) -to-inosine (I) / (G) base conversion by Adenosine deaminases acting on RNA (ADARs) is a significant epigenetic mechanism (**Fig. 4**) [47]. The target of ADARs is double-stranded RNA produced both intra- and intermolecularly. The hydrolytic deamination occurred at position C6 which converts adenosine (A) to inosine (I) [45, 48].

So far, A-to-I conversion in mammals, *Drosophila melanogaster*, *Caenorhabditis elegans*, and squid has been observed only in the nervous system. In *Drosophila*, genetic inactivation of ADAR results in severe behavioral impairments and neurological symptoms, including paralysis, poor locomotion, and tremors that worsen with age [49]. In contrast to worms and flies, mammals are entirely reliant on ADARs. A single ADAR1 allele causes death by embryonic day 14.5 in mice, as well as abnormalities in blood cell proliferation and differentiation [50]. With the deletion of the ADAR2 allele in mice, they die within 20 days after birth and become progressively seizure-prone after 12 days [51].
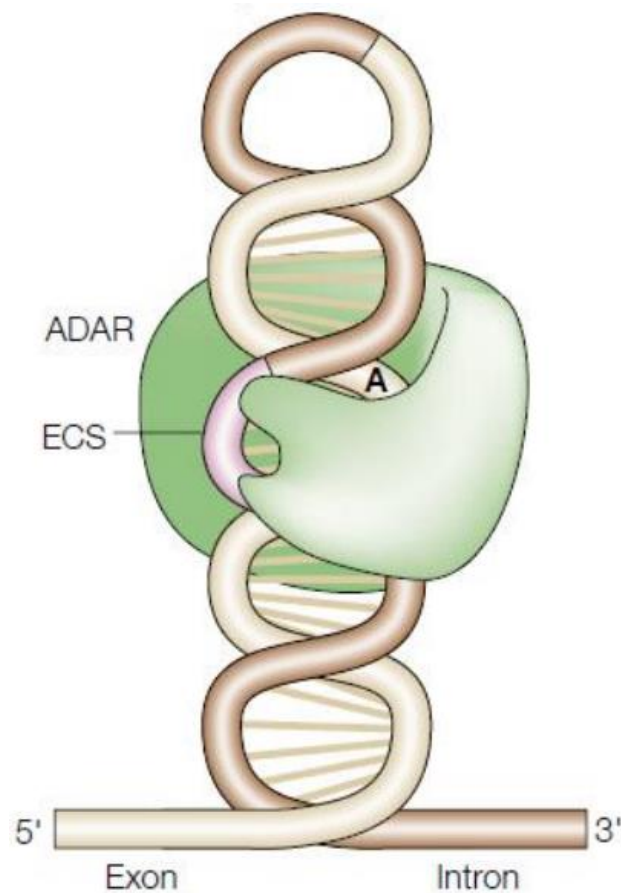
Fig 4: ADARs recognize the double strands RNA between the editing site and the complementary editing site sequence ECS. Normally, ECS is located in the downstream of the

intron. The ADAR binds to the double-stranded RNA (dsRNA) via its dsRNA binding domains (dsRBDs) and deaminates specific adenosines to inosines [27].

On the other hand, editing of certain adenosines in a pre-mRNA can be detrimental, such as in the glutamine/arginine (Q/R) site in GluR-B [52]. In earlier studies, it was demonstrated that the Ca^{2+} permeability of heteromeric -amino-3-hydroxy-5-methylisoxazole-4-propionate (AMPA) receptors, which regulate fast-excitatory-synaptic transmission in the central nervous system, is controlled by editing at this site in the Glu-R transcript [53]. Mice unable to edit exclusively at this site experienced enhanced Ca^{2+} permeability of AMPA receptors, leading to epileptic episodes, and perished three weeks after birth [53].

1.6 Mechanism of U to C editing

In higher eukaryotes, RNA editing commonly involves A-to-I and C-to-U conversions. However, in plant species, U-to-C RNA editing also occurs [54-56]. In *arthropods*, U-to-C RNA editing has been observed in the generation of tissue-specific calcium channels [57]. RNA editing can affect many important biological processes such as alternative splicing and microRNA processing [54, 58-60]. In 12-day-old *Arabidopsis* seedlings, the PPR gene AT2G19280 was discovered to be a target of U-to-C RNA editing by my colleague [61]. By comparing to the public RNA-seq databases, we identified seven U-to-C RNA editing sites in *Arabidopsis*: AT2G16586, AT5G42320, AT5G02670, AT3G41768, AT4G32430, AT3G47965, and AT5G52530 [62].

1.7 APOBEC family and its function

The discovery of a C-to-U base change in the apolipoprotein B (apoB) mRNA marked the beginning of the APOBEC field. RNA editing enzyme of APOBEC1 was characterized as the first real member of the AID/APOBEC family [63]. As we know, the human APOBEC family consists of 11 major gene products, including APOBEC1 (A1), activation-induced deaminase (AID), APOBEC2 (A2), APOBEC3A-H (A3A-H), and APOBEC4 (A4) (**Fig. 5, Table 1**) [64].
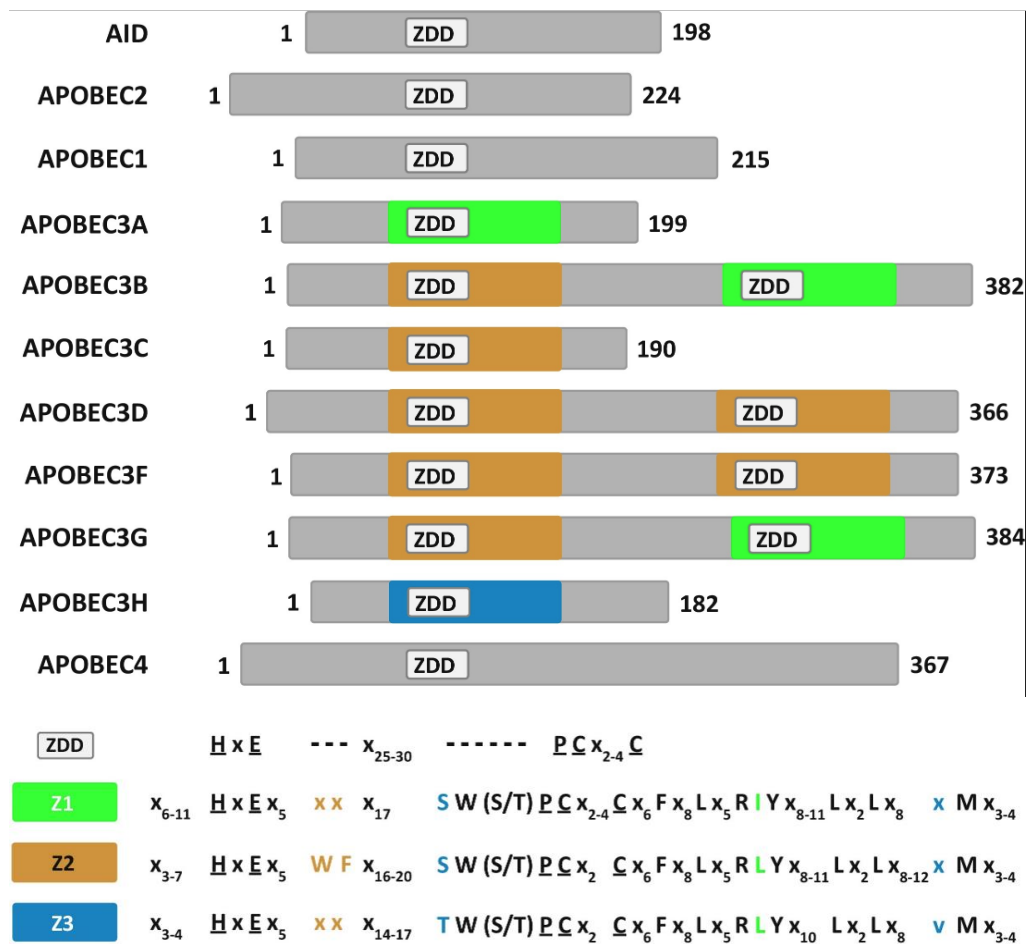
Fig 5: Humans express several APOBEC proteins, including AID, A2, A1, A3A - A3H, and A4. Such proteins can be divided into three paralogs by their different cytidine deaminase domain, especially of the zinc-dependent deaminase (ZDD) motif: Z1 (green), Z2 (orange), and Z3 (blue) [62].

The diagram displays the proteins' gene duplication and likely divergence, with their overall lengths given at the C terminus and aligned to the first ZDD motif. The protein lengths are shown to scale. The typical amino acid pattern for the ZDD, Z1-, Z2-, and Z3-type cytidine deaminase domains are shown at the bottom of the panel.

The APOBEC3 (A3) family of cytidine deaminases, plays an important role in the inherent immune system of vertebrate by limiting endogenous retroelements and foreign viruses [66-67]. Recent studies believed that the enzyme of A3 may also aid retroviruses in evading medication treatment and identification by adaptive immune systems [69, 70]. In primates, the A3 family includes seven homologous enzymes, which contain HX1EX23-24CX2-4C patterns and either one (APOBEC3A, APOBEC3C, and APOBEC3H) or two (APOBEC3B, APOBEC3D-E, APOBEC3F, and APOBEC3G) zinc (Zn)-coordinating catalytic domains (X represents any amino acid) [71-73]. During the deaminase reaction, the glutamic acid residue can serve as a proton shuttle, while the histidine and cysteine residues coordinate Zn^{2+} [74]. Among the APOBECs, APOBEC2 and APOBEC4 are the least studied, primarily because neither protein's expression is associated with any distinctive symptoms [75, 76]. Although APOBEC2 cannot alter RNA or DNA, it can bind DNA with affinities that are significantly higher than any other family member [77]. Recent studies suggested that APOBEC4 may be vital in both the antiviral response in birds and promoter regulation within mammalian cells [78, 79].

Protein	Substrate				Subcellular localization	Tissue expression	Target sequence	Known biological functions
	Host		Viral					
	RNA	DNA	RNA	DNA				
Generalists								
A3A	ss	ss	ss	ss	N/C	Monocytes and macrophages, bone marrow and lymphoid tissue, urinary tract and bladder	5′ -TC-3′	Protection from viruses
A3G	ss	No	ss	ss	C	Peripheral blood cells, IFNα-activated cells, bone marrow and lymphoid tissue, breast tissue, reproductive system, urinary tract and bladder, gastrointestinal tract, liver and gallbladder, respiratory tract	5′ -CC-3′	Protection from viruses and retroelements
APOBEC1	ss	ss	ss	No	N/C	Gastrointestinal tract (mouse and human); immune cells (mouse only)	5′ -AC-3′	Regulation of cholesterol metabolism (mouse and human); modulation of monocyte transcriptome (mouse only); potentially antiviral
Specialists								
AID	No	ssDNA and RNA/DNA hybrid	No	No	N/C	Germinal centre B cells, lymphoid tissue	5′ -WRC-3′ ☒	Secondary antibody diversification
APOBEC2	No	ss/ds binding only	No	No	N/C	Muscle (skeletal and cardiac); some B cell/T cell subsets	ND	ND
Unassigned								
A3B	ND	ss	No	ss	N	IFNα-activated liver cells, bone marrow, gastrointestinal tract, kidney, urinary tract and bladder	5′ -TC-3′	Protection from viruses
A3C	No	ND	No	ss	N/C	Peripheral blood cells, bone marrow and lymphoid tissue, skin, muscle, breast tissue, reproductive system, kidney, urinary tract and bladder, gastrointestinal tract, liver and gallbladder, proximal digestive tract, respiratory tract, endocrine tissues	5′ -TC-3′	Protection from viruses
A3D	No	No	No	ss	C	Peripheral blood cells, bone marrow and lymphoid tissue, gastrointestinal tract, female reproductive system	5′ -TC-3′	Protection from viruses
A3F	No	No	No	ss	C	Peripheral blood cells, IFNα-activated cells, bone marrow and lymphoid tissue, reproductive system, endocrine tissues, muscle	5′ -TC-3′	Protection from viruses
A3H	ND	ND	No	ss	N/C	Peripheral blood cells, bone marrow and lymphoid tissue.	5′ -TC-3′	Protection from viruses

Table 1: AID/APOBEC members are categorized as specialists or generalists depending on the flexibility of their substrate and functional limitations [75].

1.8 ADAR family and its function

ADARs are modular proteins that contain different domains for distinct functions [45, 46] (**Fig. 6**). In mammals, ADAR family includes four members with varying biological activity or cellular localization [49]. Highest expression levels ADAR1 and ADAR2 were found in the central nervous system, which are involved in a lot of post-

transcriptional editing [50, 51]. ADAR3 only expressed in some certain regions of the brain, but it is catalytically inactive. Until now, the exact function of ADAR3 has not been explained [80]. The testis-expressed nuclear RNA-binding protein (TENR) also belong to the member of the family, which only expressed in the testis [81].

ADARs is capable to deaminate adenosine to inosine non-specifically in long double-stranded RNA (dsRNA) or site-specifically in transcripts [47]. During the dsRNA stage of a virus, non-specific adenosine deamination can disrupt the open reading frame of the viral genome, thereby exerting a defense effect on the virus [82]. Another role of ADARs is site-specific deamination of a single adenosine, which can alter the codons of the open reading frame, including the start or stop codon, or affect splicing or untranslated regions [26, 52, 53].
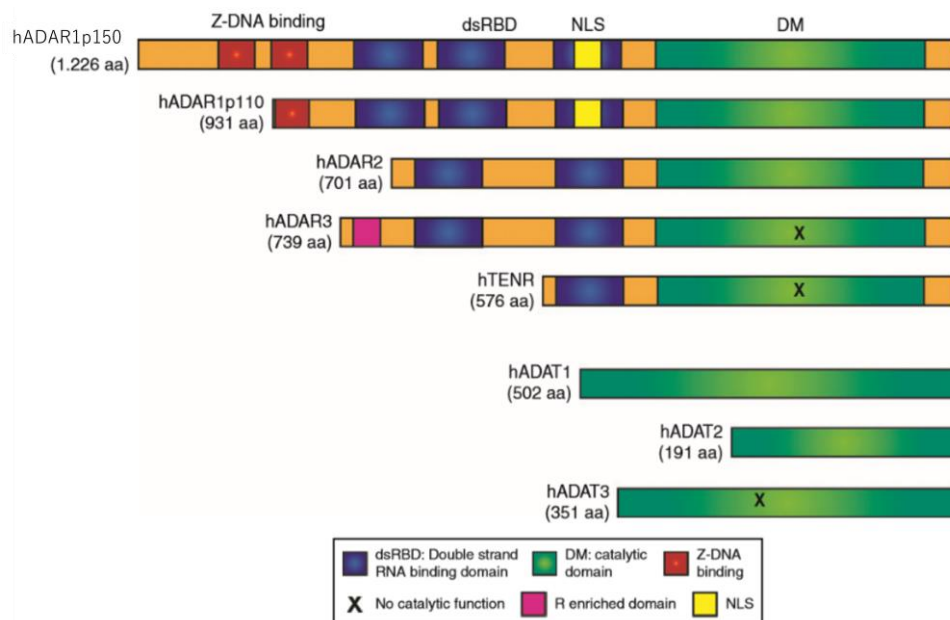
Fig. 6: The catalytic deaminase domain and dsRNA-binding domains (blue) comprise the protein domains of ADARs, along with the DM domain (green box). The DM domain of ADAT1, ADAT2, and ADAT3 also binds to tRNA. The DM domain is marked with an (X) for ADAR3, TENR, and

ADAT3, indicating their inactivity. ADAR1's N-terminal domain binds to Z-DNA, while ADAR3 has a region enriched with arginine (R) residues in that domain [83].

ADAR1 is expressed in two isoforms due to alternative splicing. The full-length ADAR1 isoform, with a molecular weight of 150-kDa, is present in both the cytoplasm and nucleus of human cells, while the shorter 110-kDa isoform is exclusively located in the nucleus [84].

1.9 Strategies for site-directed RNA editing with ADAR

1.9.1 SNAP-tag system

Initially, fusion proteins were labeled in vivo using the *SNAP-tag* technique, which allows for the covalent labeling of proteins with a small molecule. As a human DNA repair protein O6-alkylguanine-DNA alkyl transferase (hAGT) catalyzes the transfer of an alkyl group from O6-alkylguanine to a cysteine residue, which is responsible for the system's covalent tagging (**Fig. 7**) [37, 85].

In Stafforst group, they adapted this system by fusing a SNAP-tag to the N-terminus of ADAR's deaminase domain (DD) and linking it to the guide RNA (gRNA) (**Fig 8**) [86]. In this system, a 17-nucleotide-long guide RNA was found to be the most suitable for matching antisense-oligos (AOs) that contain a single C mismatch under the target A and are positioned in the middle of the guide RNA [87]. Additionally, they used 2'-O-methyl modification of the guide RNA to increase the RNA's resistance to nucleases [37, 87].
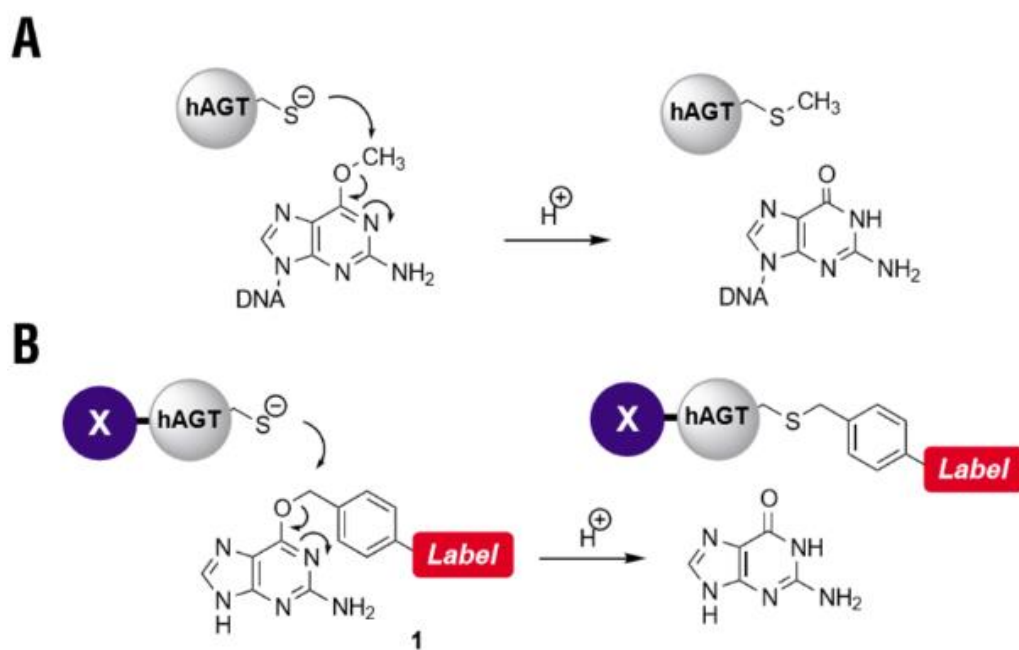
Fig. 7: (A) DNA repair mechanism by hAGT. (B) O6-benzylguanine (BG) derivatives are used to covalently mark an X-hAGT fusion protein [85]
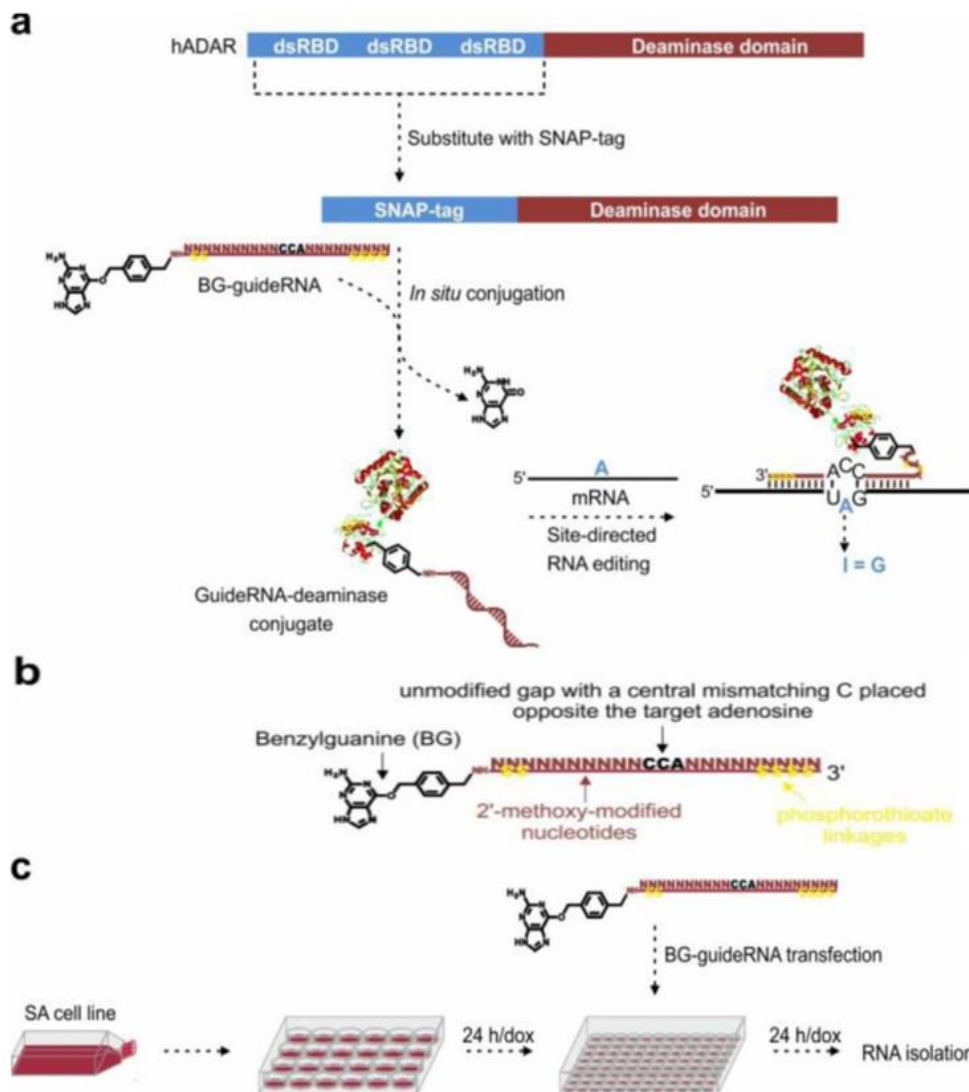
Fig. 8: (a) The deaminase domain of hADAR (amino acid 798–1226) is replaced by the SNAP-tag. Together with the target RNA, the guide RNA can form a secondary structure essential for the A-to-I deamination reaction. (b) A typical 5'-CCA anticodon benzylguanine-guide RNA that targets a UAG site. (c) Experimental setup: Cells stably integrated with SNAP-ADAR are seeded onto 24-well plates with doxycycline-containing media to induce SNAP-ADAR expression. The guide RNA is reverse-transfected into the cells 24 hours later. After 24 hours, the cells are lysed for RNA isolation to study RNA editing [88].

Sanger sequencing and fluorescence microscopy were used to evaluate editing. The editing efficiency was reported by the authors to be around 30% under these experimental circumstances [87]. Additional testing of the system revealed that, depending on the version of the DD utilized, the SNAP-tag DD can also edit UAG triplets in endogenously expressed targets in HEK-293 cells (GAPDH, ACTB, for example) with efficiency up to 90% [88].

1.9.2 λN-BoxB system

A synthetic peptide was utilized in the *λN-BoxB* system. Comparing with the full-length protein, the synthetic peptide consisting of the 22 N-terminal amino acids of the lambda N-protein, has a similar affinity and specificity for combining with the BoxB RNA element (**Fig. 9**) [37, 89]. BoxB is a 15-bp RNA element located downstream of nutR and nutL sites that can fold into a hairpin structure [90].
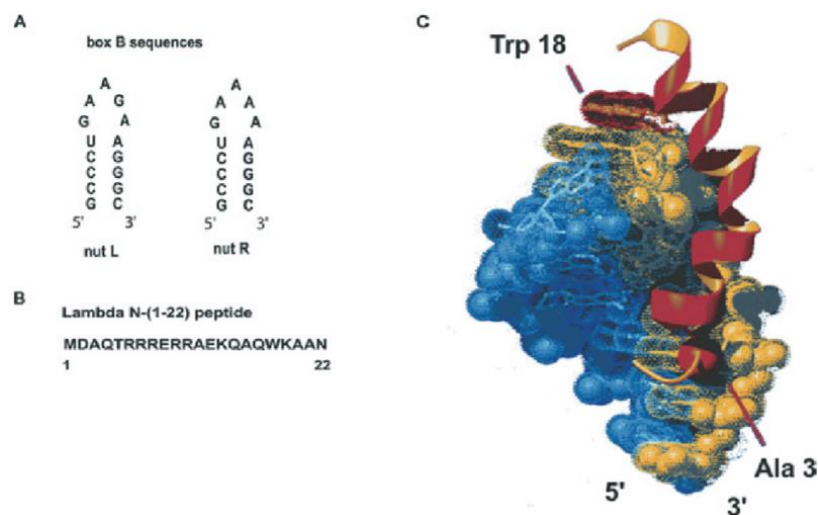
Fig. 9: BoxB RNA and lambda N peptide.   (A) The sequence of lambda *nutL* and *nutR.* (B) the 1–22 peptide of the lambda N protein sequence. (C) Illustration of the relationship between the lambda N peptide 1-22 and the nutL stem-loop. The RNA is displayed as a surface with dots, The

spheres for the C3', C5', O3', O5', O1P, O2P, and P atoms depict the phosphate backbone. The

smoothed ribbon represents the backbone of the C peptide [91].

The Gonzalez group conducted RNA editing research based on the *λN-BoxB* system. The results showed that the system was able to amend a premature termination codon in mRNA encoding the cystic fibrosis transmembrane conductance regulator anion channel at the *in vitro*. A signal of enhanced green fluorescent protein (EGFP) was observed in the successfully edited HEK-293T cells (**Fig. 10**) [37, 92].
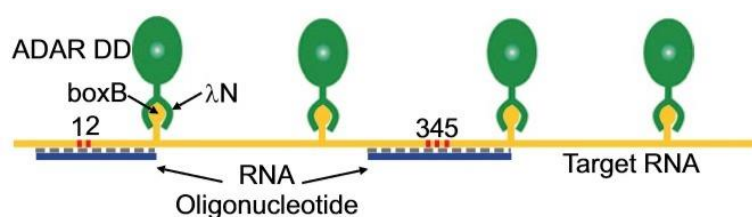
Fig. 10: The λN-BoxB system used in the Gonzalez group research. a diagram illustrating the experimental strategy. RT-PCR demonstrated editing at five adenosines after treating 4boxB target RNA with N-DD and a complementary antisense RNA oligonucleotide (1–5, in red) [92].

After confirming that the *λN-BoxB* system could correct mutations, the Gonzalez group attempted to increase its editing efficiency by using multiple copies of the boxB and *λN* elements. They found that the system containing two boxB elements fused with the guide RNA and four tandem repeats of the *λN* unit was the most effective (**Fig. 10**). [37, 93]

They transfected mCherry-EGFP containing a premature termination codon (W58X) into HEK-293T cells and confirmed the editing efficiency by both fluorescence imaging

and Sanger sequencing. The results showed that the correction ratio was approximately 20% to 70% around the target but highly dependent on the context (**Fig. 11**) [37, 92, 93].
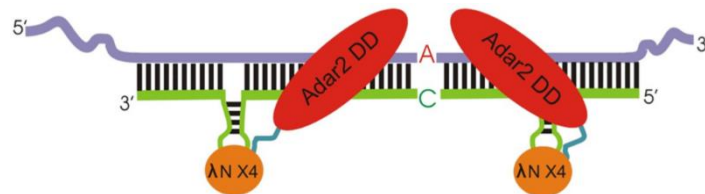
Fig. 11: *λN-BoxB* strategy. Fused to the ADAR2 deaminase domain (red color), the guide RNA introduces the editing enzyme with N peptides (λN x4, orange color) to the target mRNA (green color). The guide RNA consists of three regions: i) two BoxB RNA hairpins positioned strategically to promote efficient editing. ii) a central region fully complementary to the target mRNA except for a C mismatch at the target. iii) two fully complementary regions at each terminus of the gRNA to stabilize it [37].

1.9.3 CRISPR-Cas13b (REPAIR)

CRISPR-associated (Cas) proteins and prokaryotic clustered regularly interspaced short palindromic repeats (CRISPR) RNAs function as an adaptive immune system to defend bacteria and archaea against invading genetic material such as bacteriophages and plasmids [94]. Cas13 is a type VI CRISPR-associated RNase which can further divided into four subtypes: Cas13a, Cas13b, Cas13c, and Cas13d on the basis of the phylogenetic relationship of their effector complexes (**Fig. 12**) [37, 95-97]. Two Higher Eukaryotes and Prokaryotes Nucleotide-binding (HEPN) domains which mediate RNA cleavage were included in each of the subtypes.
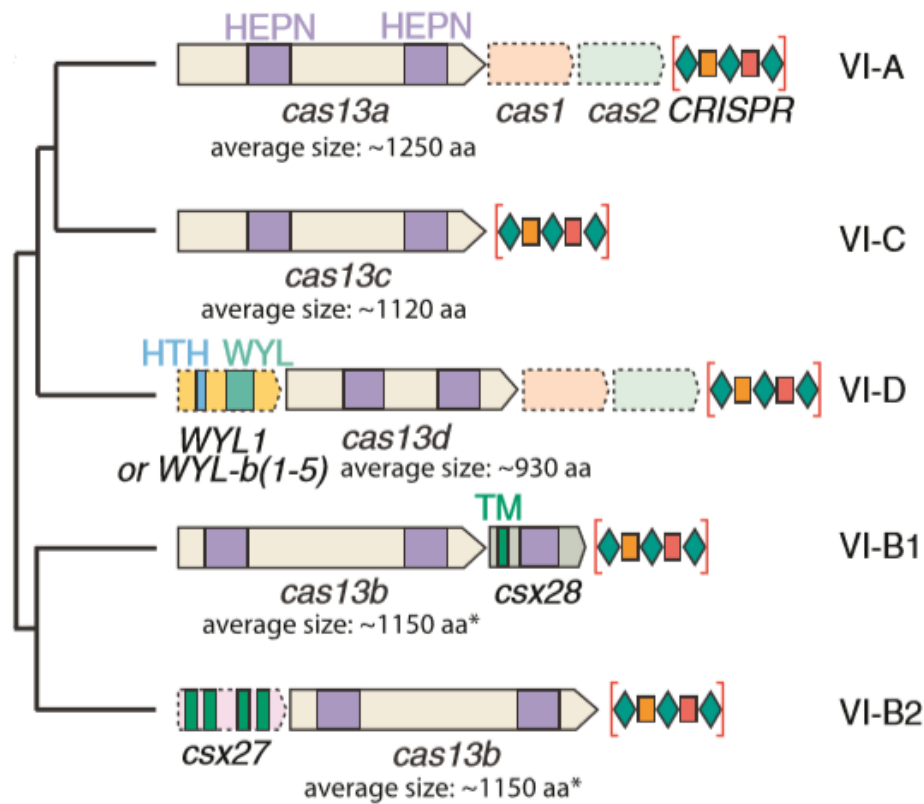
Fig. 12: The figure illustrates the genomic locus layouts of different Type VI CRISPR subtypes. The phylogenetic tree represents the relationships between the subtypes. The Cas13 genes are labeled and roughly to scale. The average size of each subtype's Cas13 protein is indicated (* Cas13b size includes both VI-B1 and VI-B2 subtypes). The genes indicated by dashed lines are not always present in the subtype. Abbreviations of HEPN stands for Higher Eukaryotes and Prokaryotes Nucleotide-binding domain, HTH for helix-turn-helix domain, WYL for WYL domain, and TM for the projected transmembrane-spanning region. CRISPR arrays are represented by green diamonds, while spacer sequences are depicted by orange and red rectangles [97].

To assess CRISPR-Cas13b activity in mammalian cells, the authors fused dCas13b with the deaminase domain of human ADAR1 or ADAR2, and then transfected this construct into HEK-293FT cells with Cluc (an RNA-editing reporter). Meanwhile, a

nonsense mutation was introduced (W85X(UGG-UAG)) (**Fig. 13**). Editing efficiency

was assessed by monitoring luciferase activity and Illumina sequencing after 48 hours of

the transfection. The editing efficiency of up to 30% were observed for all the targets [37,

98]. This new system was given a name of REPAIR (RNA Editing for Programmable A

to I Replacements).

Fig. 13: A) The diagram illustrates the fusion of dCas13b with the deaminase domain of human

ADAR (ADARDD) for RNA editing. ADARDD deaminates adenosine to inosine in dsRNA, and

when joined to catalytically dead Cas13b (dCas13b), it can edit RNA. B) A schematic of the target

design for the Cypridina luciferase W85X target is shown. The stop codon in the wildtype tryptophan

is restored by deamination of the target adenosine. The number of bases between the mismatched

cytidine and the spacer's 3' end is known as the mismatch distance [98].

1.9.4 AI-REWIRE

Pumilio and *fem*-3 mRNA-binding factors (PUF) homology proteins combine with

specific RNA sequences, and then function in regulating the translation and stability of

target mRNAs (**Fig. 14**). In 2001, Traci M. TanakaHall's group used RNA selection

experiments to identify the Homo sapiens Pumilio 1 homology domain (HsPUM1-HD)
[99].
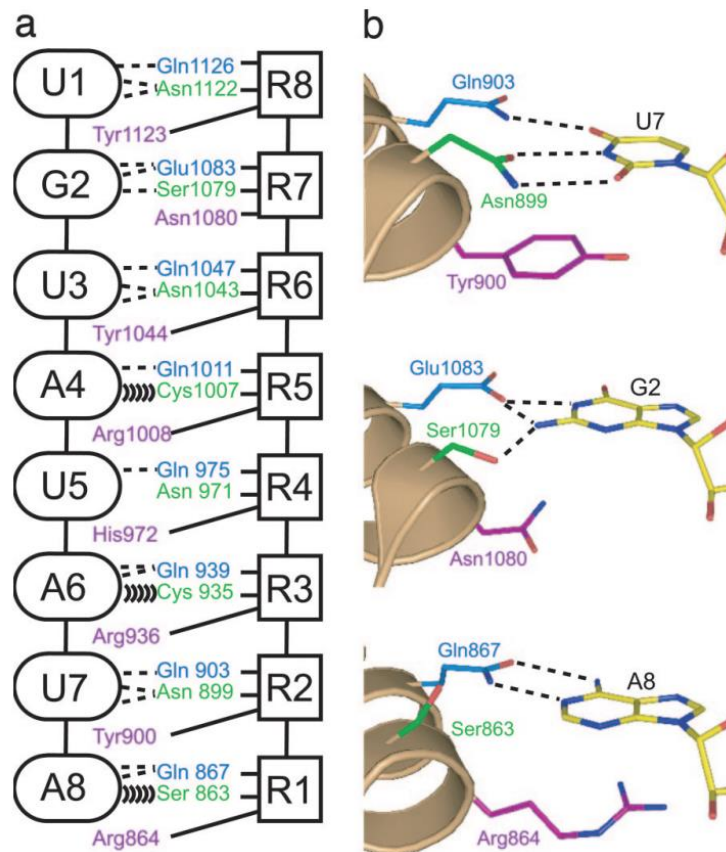
Fig. 14: Recognition of RNA by HsPUM1-HD. (i) The figure shows a schematic representation
of the RNA:protein interaction between HsPUM1-HD and NRE RNA. The RNA bases are
represented by ovals, and the protein repeats are represented by squares. Dashed lines indicate
hydrogen bonds, while parentheses indicate van der Waals contacts. Each repeat has the same
positions for the blue, green, and purple side chains. (ii) The figure demonstrates the interaction of
HsPUM1-HD with adenine, guanine, or uracil at the top, middle, or bottom of the RNA sequence.
Stick representations are used to display the RNA and side chains that interact with it. The nitrogen
atoms are represented in dark blue, oxygen atoms in red, and carbon atoms in yellow, light blue,
green, or purple. The PDB ID for this structure is 1M8Y [99, 100].

The Pumilio (PUM) repeats of PUF proteins are modular and RNA sequence specific due to their recognition code for nucleotides A, U, and G [100]. In 2011, modified splicing factors and PUF domains were created to regulate alternative splicing and detect targets with multiple cytosines using the C-recognition code [101].

Recently, in the Wang group, a new system was developed, which was called REWIRE (RNA Editing with Individual RNA-binding Enzyme). It performed precise base editing by a single engineered protein, without the need for guide RNA (as shown in Fig. 15). In HEK 293T cells, they used AI-REWIRE to perform genetic code restoration, and the ADAR1 E1008Q enzyme showed an editing efficiency of approximately 50%, while ADAR2 E488Q showed an efficiency of approximately 80% [102].
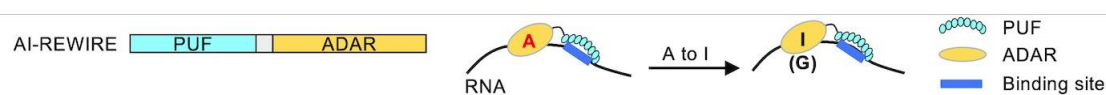
Fig. 15: Specifically designing AI-REWIREs for A to I editing. AI-domain REWIRE's setup (left); a schematic of the RNAs' particular A-to-I base editing caused by AI-REWIRE (right). The AI-REWIRE can induce A-to-I RNA editing on the single-strand RNA.**[102].**

1.10 Strategies for site-directed C-to-U RNA editing

1.10.1 RESCUE

In 2019, the Zhang group utilized mutated ADAR2 to perform site-directed cytidine to uridine RNA editing [104]. It was named as RESCUE (RNA Editing for Specific C-to-

U Exchange) (**Fig. 16**). They discovered that V351G and K350I mutations in the ADAR2 catalytic core could induce C-to-U deamination reaction.
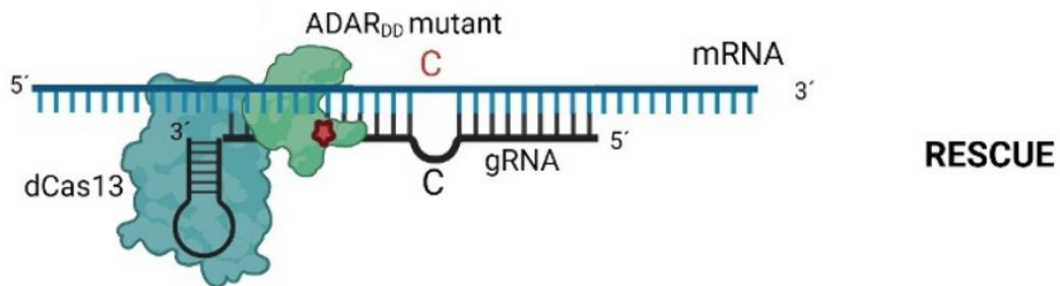
Fig. 16: Schematic of RESCUE system. This schematic showed the dCas13 protein was linked to the mutated ADAR2 deaminase domain. The target cytidine was marked as red on the target mRNA [38].

It was noteworthy, in addition to producing ~80% C-to-U editing efficiency on the target RNA, RESCUE system can also produce A-to-I deamination editing [38].

1.10.2 CURE

In 2020, the Chi group developed the CURE system, a gRNA-based approach that uses human APOBEC3A fused with dCas13b for C-to-U RNA editing (**Fig 17**) [38, 105]. To induce editing, they designed special gRNAs that can induce the formation of loops at the target sites. The advantage of using human APOBEC3A is that it does not produce A-to-I deamination, thus eliminating the off-target effects associated with the RESCUE system.
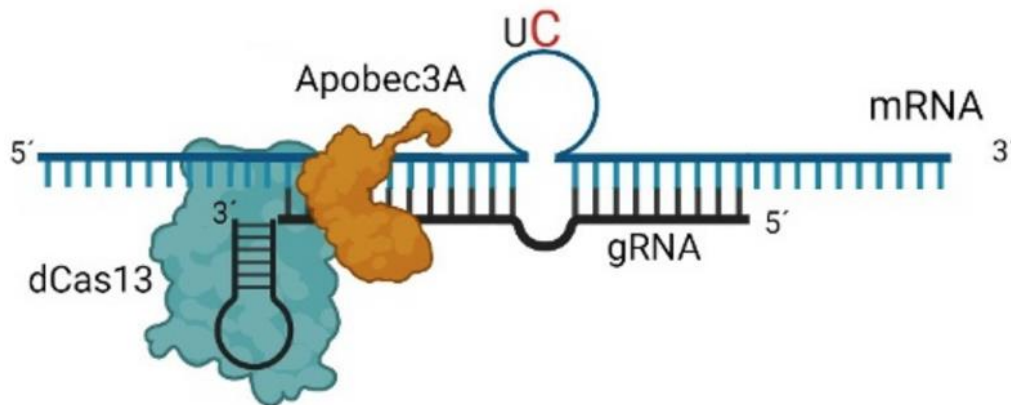
Fig. 17: Schematic of CURE system. They used the dCas13 protein linked to the human

APOBEC3A. The guide RNA can induce nucleotide loop on the target mRNA **[38].**

The CURE system has shown an on-target editing efficiency of approximately 60%. However, a limitation of this system is that the edited base U must be preceded by the editing target C in order to perform deamination editing. Additionally, there is a concern that free APOBEC3A may lead to unintended C-to-T mutations in DNA, which could be addressed through optimization of the amino acid sequence of APOBEC3A [38, 106].

1.10.3 CU-REWIRE

CU-REWIRE is a system developed by the Wang group for C-to-U RNA editing (**Fig 18**). In this system, APOBEC3A is fused to the PUF protein using the XTEN linker. This system can achieve editing efficiencies of up to 60%. A significant advantage of this system is that it does not require a specific sequence context for editing, allowing for more flexible editing of C residues.
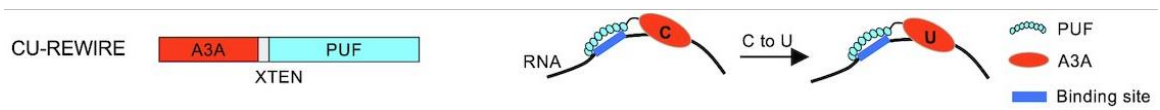
Fig. 18: Schematic of CU-REWIRE system. There is XTEN linker between the APOBEC3A

deaminase domain and the PUF protein. [102]

However, like the CURE system, CU-REWIRE requires the target U to be preceded by the editing target C for deamination editing to occur. The Wang group also performed the R-loop assay and found that CU-REWIRE can induce deamination on DNA, which poses a potential risk in practical applications.

1.11 MS2 RNA-protein interactions

The three-hybrid system originated from the yeast two-hybrid system [107] and was developed to analyze the interactions between RNA-and protein, for example, by measuring the ability to grow or metabolize chromogenic compounds (**Fig. 19**) [108].
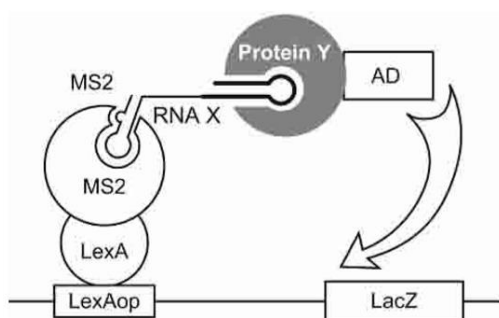
Fig 19: The three-hybrid system is a modified version of the yeast two-hybrid system [100], used to detect RNA-protein interactions based on growth or the ability to metabolize chromogenic compounds [108]. This system uses RNA and protein building blocks and strains L40coat and YBZ-1, both of which contain LacZ and HIS3 and are operated by LexA operators [109]**.**

In Fig. 19, the protein comprises a fusion of LexA/MS2 coat protein. An RNA containing MS2-binding sites binds to the MS2 protein (**Fig 20**). The RNA also contains a sequence of interest, X, which binds to an RNA-binding polypeptide, Y [109]. The activation domain and Y are linked (AD). The reporter gene will be activated when the required interactions occur.

The MS2 coat protein acts as a host for the replicase mRNA's translation and as a structural protein for the bacteriophage particle. Precise binding of the coat protein to the start codon leads to translational inhibition (AUG) [110, 111].

The X-ray structure and the crystallographic data indicate that the MS2 proteins form and assembled as dimer orientated in antiparallel orientation (**Fig 20**) [112].
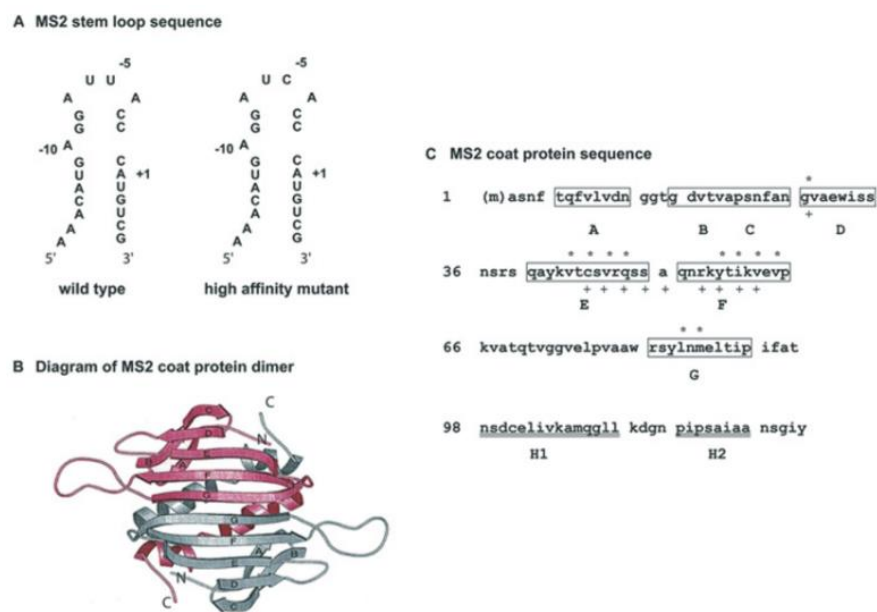
Fig 20: The MS2 coat protein and the RNA stem-loop are illustrated. (A) The MS2 stem-loop is depicted in its wild-type and high-affinity C-loop forms. The first base of the numbering system is the replicase mRNA's start codon, AUG. (B) The interaction between two monomers of the MS2 coat protein is shown, while the N- and C-termini of each monomer labeled, and the beta-sheets numbered

from A to G. (C) The amino acid sequence of the MS2 coat protein is shown in, with the numbers corresponding to the processed protein from which the original methionine has been removed. The amino acids are boxed, and their corresponding beta-sheets (from A to G) are shown in bold. The helical segments are labeled H1 and H2 and are double-underlined. The symbols * or + denote the RNA-binding amino acids of the upper or lower subunits of the dimer, respectively [112].

1.12 Aim of the study

In the context of gene therapy, the target is often the entire gene or transcript, but my objective is to correct a single nucleotide mutation that leads to premature stop codons or non-functional proteins, such as in the case of disease including a G to A or T to C mutation. The correction of these errors through base substitution can restore the correct sense codon and ultimately lead to functional protein expression. The efficacy of this correction and the development of an artificial RNA editing system for therapeutic purposes is crucial. To achieve this goal, it is essential to increase the editing efficiency of the MS2 RNA editing system.

Therefore, the aims of this study are:

1. Improve the editing efficiency of an artificial deaminase complex by ADAR1, APOBEC3A and APOBEC3G link with MS2 system, in order to restore the G-to-A (TAG amber stop codon) as well as T-to-C (CAC-TAC, BFP to GFP) mutation.

2. Improve the overall editing efficiency of this system.

1.13 Reference

[1] "Genetic Engineering". Genome.gov. Retrieved 20 February 2022.

[2] P. Byrne. Genetically Modified (GM) Crops: Techniques and Applications. Crop Series. Fact Sheet No. 0.710.

[3] Maeder ML, Gersbach CA. Genome-editing Technologies for Gene and Cell Therapy. Mol Ther. 2016 Mar;24(3):430-46.

[4] Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature. 2016 May 19;533(7603):420-4.

[5] Bak, Rasmus O.; Gomez-Ospina, Natalia; Porteus, Matthew H. (August 2018). "Gene Editing on Center Stage". Trends in Genetics. 34 (8): 600–611.

[6] Goodman MF, Hopkins R, Gore WC. 2-Aminopurine-induced mutagenesis in T4 bacteriophage: a model relating mutation frequency to 2-aminopurine incorporation in DNA. Proc Natl Acad Sci U S A. 1977 Nov;74(11):4806-10.

[7] Woolf TM (April 1998). "Therapeutic repair of mutated nucleic acid sequences". Nature Biotechnology. 16 (4): 341–4.

[8] Jasin M (June 1996). "Genetic manipulation of genomes with rare-cutting endonucleases". Trends in Genetics. 12 (6): 224–8.

[9] Moore JK, Haber JE (May 1996). "Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in Saccharomyces cerevisiae". Molecular and Cellular Biology. 16 (5): 2164–73.

[10] Boulton SJ, Jackson SP (September 1996). "Saccharomyces cerevisiae Ku70 potentiates illegitimate DNA double-strand break repair and serves as a barrier to error-prone DNA repair pathways". EMBO J. 15 (18): 5093–103.

[11] Wilson TE, Lieber MR (1999). "Efficient processing of DNA ends during yeast nonhomologous end joining. Evidence for a DNA polymerase beta (Pol4)-dependent pathway". J. Biol. Chem. 274 (33): 23599–23609.

[12] Budman J, Chu G (Feb 2005). "Processing of DNA for nonhomologous end-joining by cell-free extract". EMBO J. 24 (4): 849–60.

[13] Espejel S, Franco S, Rodríguez-Perales S, Bouffler SD, Cigudosa JC, Blasco MA (May 2002). "Mammalian Ku86 mediates chromosomal fusions and apoptosis caused by critically short telomeres". The EMBO Journal. 21 (9): 2207–19.

[14] Guirouilh-Barbat J, Huck S, Bertrand P, et al. (June 2004). "Impact of the KU80 pathway on NHEJ-induced genome rearrangements in mammalian cells". Mol. Cell. 14 (5): 611–23.

[15] McVey M, Lee SE (November 2008). "MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings". Trends Genet. 24 (11): 529–38.

[16] Malzahn, Aimee; Lowder, Levi; Qi, Yiping (2017-04-24). "Plant genome editing with TALEN and CRISPR". Cell & Bioscience. 7 (1): 21.

[17] Pardo, B; Gomez-Gonzales, B; Aguilera, A (March 2009). "DNA repair in mammalian cells: DNA double-strand break repair: how to fix a broken relationship". Cellular and Molecular Life Sciences. 66 (6): 1039–1056.

[18] Bolderson, Emma; Richard, Derek J.; Zhou, Bin-Bing S. (2009). "Recent Advances in Cancer Therapy Targeting Proteins Involved in DNA Double-Strand Break Repair". Clinical Cancer Research. 15 (20): 6314–6320.

[19] Coïc E, Feldman T, Landman AS, Haber JE (2008). "Mechanisms of Rad52-independent spontaneous and UV-induced mitotic recombination in Saccharomyces cerevisiae". Genetics. 179 (1): 199–211.

[20] Goldfarb T, Lichten M (2010). "Frequent and efficient use of the sister chromatid for DNA double-strand break repair during budding yeast meiosis". PLOS Biology. 8 (10): e1000520.

[21] Stringer JM, Winship A, Zerafa N, Wakefield M, Hutt K. Oocytes can efficiently repair DNA double-strand breaks to restore genetic integrity and protect offspring health. Proc Natl Acad Sci U S A. 2020 May 26;117(21):11513-11522.

[22] Sands MS, Bogenhagen DF. Two zinc finger proteins from Xenopus laevis bind the same region of 5S RNA but with different nuclease protection patterns. Nucleic Acids Res. 1991 Apr 25;19(8):1797-803.

[23] Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science. 2012 Aug 17;337(6096):816-21.

[24] Peters DT, Cowan CA, Musunuru K. Genome editing in human pluripotent stem cells. 2013 Apr 29. In: StemBook [Internet]. Cambridge (MA): Harvard Stem Cell Institute; 2008–.

[25] Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. Cell 157(6), 1262–1278.

[26] Thomas, G., Shannon, J. S., Sai-lan, S. & Jia, L. Genome editing technologies: Principles and applications. *Cold Spring Harbor Perspect. Biol.* **8**, a023754. (2016)

[27] Keegan LP, Gallo A, O'Connell M a. The many roles of an RNA editor. Nat Rev Genet. 2001; 2(11): 869–78.

[28] Maas S, Rich A. Changing genetic information through RNA editing. BioEssays 2000; 22: 790–802.

[29] Benne R, Van Den Burg J, Brakenhoff JPJ, Sloof P, Van Boom JH, Tromp MC. Major transcript of the frameshifted cox II gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. Cell 1986; 46: 819–826.

[30] Powell LM, Wallis SC, Pease RJ, Edwards YH, Knott TJ, Scott J. A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine. Cell 1987; 50: 831–840.

[31] Blanc V., and Nicholas O. Davidson. C-to-U RNA Editing: Mechanisms Leading to Genetic Diversity. The Journal of Biological Chemistry, 278(3): 1395–1398, (2003).

[32] Gott JM, Emeson RB. Functions and mechanisms of RNA editing. Annu Rev Genet. 2000;34:499-531.

[33] Ruchika, Okudaira C, Sakari M, Tsukahara T. Genome-Wide Identification of U-To-C RNA Editing Events for Nuclear Genes in Arabidopsis thaliana. Cells. 2021 Mar 12;10(3):635.

[34] Ruchika, Tsukahara T. The U-to-C RNA editing affects the mRNA stability of nuclear genes in Arabidopsis thaliana. Biochem Biophys Res Commun. 2021 Sep 24;571:110-117.

[35] Maas S, Rich A. Changing genetic information through RNA editing. BioEssays **2000**; *22*: 790–802.

[36] Paul Vogel; Thorsten Stafforst. Critical Review on Engineering Deaminases for Site-Directed RNA Editing. *Curr. Opin. Biotechnol*. **2019**, *55*, 74–80.

[37] Montiel-Gonzalez; M.F. Quiroz; J.F.D. Rosenthal; J.J. Current Strategies for Site-Directed RNA Editing Using ADARs. Methods **2019**, *156*, 16–24.

[38] Genghao Chen; Dhruva Katrekar; Prashant Mali. RNA-Guided Adenosine Deaminases: Advances and Challenges for Therapeutic RNA Editing. Biochemistry **2019**, *58*, 1947–1957.

[39] Prohaska KM, Bennett RP, Salter JD et al (2014) The multifaceted roles of RNA binding in APOBEC cytidine deaminase functions. Wiley Interdiscip Rev RNA 5:493

[40] Sharma S, Patnaik SK, Thomas Taggart R et al (2015) APOBEC3A cytidine deaminase induces RNA editing in monocytes and macrophages. Nat Commun 6:6881.

[41] Sharma S, Patnaik SK, Taggart RT et al (2016) The double-domain cytidine deaminase APOBEC3G is a cellular site-specific RNA editing enzyme. Sci Rep 6:1–12

[42] Sharma S, Wang J, Alqassim E et al (2019) Mitochondrial hypoxic stress induces widespread RNA editing by APOBEC3G in natural killer cells. Genome Biol 20:37

[43] Chen SH, Habib G, Yang CY et al (1987) Apolipoprotein B-48 is the product of a messenger RNA with an organ-specific in-frame stop codon. Science 238:363–366

[44] Johnson DF, Poksay KS, Innerarity TL (1993) The mechanism for Apo-B mRNA editing is deamination. Biochem Biophys Res Commun 195:1204–1210.

[45] Barraud P, Allain FH. ADAR proteins: double-stranded RNA and Z-DNA binding domains. Curr Top Microbiol Immunol. 2012;353:35-60.

[46] Maas S, Rich A, Nishikura K. A-to-I RNA editing: Recent news and residual mysteries. J Biol Chem 2003; 278: 1391–1394.

[47] Bass BL. RNA editing by adenosine deaminates that act on RNA. Annu Rev Biochem 2002; 71: 817–846.

[48] Maydanovych O, Beal PA. Breaking the central dogma by RNA editing. Chem Rev 2006; 106: 3397–3411.

[49] Kazuko Nishikura. Functions and regulation of RNA editing by ADAR deaminases. Annu Rev Biochem 2010; 79: 321–349.

[50] Palladino, M. J., Keegan, L. P., O'Connell, M. A., and Reenan, R. A. A-to-I pre-mRNA editing in *Drosophila* is primarily involved in adult nervous system function and integrity.   (2000) Cell 102, 437–449

[51] Wang Q, Khillan J, Gadue P, Nishikura K. Requirement of the RNA Editing Deaminase ADAR1 Gene for Embryonic Erythropoiesis. Science 2000;290:1765–68.

[52] Higuchi M, Maas S, Single FN, Hartner J, Rozov A, et al. Point mutation in an AMPA receptor gene rescues lethality in mice deficient in the RNA-editing enzyme ADAR2. Nature 2000; 406: 78–81.

[53] Sommer, B., Köhler, M., Sprengel, R. & Seeburg, P. H. RNA editing in brain controls a determinant of ion flow in glutamate-gated channels. Cell 67, 11–19 (1991)

[54] Brusa, R. et al. Early-onset epilepsy and postnatal lethality associated with editing-deficient GluR-B allele in mice. Science 270, 1677–1680 (1995).

[55] Nagalla SR, Barry BJ, Spindel ER. Cloning of complementary DNAs encoding the amphibian bombesin-like peptides Phe8 and Leu8 phyllolitorin from Phyllomedusa sauvagei: potential role of U to C RNA editing in generating neuropeptide diversity. Mol Endocrinol 1994; 8: 943–951.

[56] Villegas J, Müller I, Arredondo J, Pinto R, Burzio LO. A putative RNA editing from U to C in a mouse mitochondrial transcript. Nucleic Acids Res 2002; 30: 1895–901.

[57] Mohini Sharma P, Bowman M, Madden SL, Rauscher III FJ, Sukumar S. RNA editing in the Wilms' tumor susceptibility gene, WTl. Genes Dev 1994; 8: 720–731.

[58] Song W, Liu Z, Tan J, Nomura Y, Dong K. RNA editing generates tissue-specific sodium channels with distinct gating properties. J Biol Chem 2004; 279: 32554–32561.

[59] Knie N, Grewe F, Fischer S, Knoop V. Reverse U-to-C editing exceeds C-to-U RNA editing in some ferns – a monilophyte-wide comparison of chloroplast and mitochondrial RNA editing suggests independent evolution of the two processes in both organelles. BMC Evol Biol 2016; 16: 1–12.

[60] Rueter SM, Dawson TR, Emeson RB. Regulation of alternative splicing by RNA editing. Nature 1999; 399: 75–80.

[61] Ota H, Sakurai M, Gupta R, Valente L, Wulff BE, Ariyoshi K et al. ADAR1 forms a complex with dicer to promote MicroRNA processing and RNA-induced gene silencing. Cell 2013; 153: 575–589.

[62] Ruchika, Tsukahara T. The U-to-C RNA editing affects the mRNA stability of nuclear genes in Arabidopsis thaliana. Biochem Biophys Res Commun. 2021 Sep 24;571:110-117.

[63] Ruchika, Okudaira C, Sakari M, Tsukahara T. Genome-Wide Identification of U-To-C RNA Editing Events for Nuclear Genes in Arabidopsis thaliana. Cells. 2021 Mar 12;10(3):635.

[64] Lyn M. Powell, Simon C. Wallis, Richard J. Pease, Yvonne H. Edwards, Timothy J. Knott, James Scott, A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine, Cell, Volume 50, Issue 6, 1987, Pages 831-840,

[65] Jason D. Salter, Ryan P. Bennett, Harold C. Smith, The APOBEC Protein Family: United by Structure, Divergent in Function, Trends in Biochemical Sciences, Volume 41, Issue 7, 2016, Pages 578-594,

[66] Pecori, R., Di Giorgio, S., Paulo Lorenzo, J. *et al.* Functions and consequences of AID/APOBEC-mediated DNA and RNA deamination. *Nat Rev Genet* **23**, 505–518 (2022).

[67] Cullen BR. 2006. Role and mechanism of action of the APOBEC3 family of antiretroviral resistance factors. Journal of Virology 80:1067–1076.

[68] Chiu YL, Greene WC. 2008. The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. Annual Review of Immunology 26:317–353.

[69] Harris RS, Dudley JP. 2015. APOBECs and virus restriction. Virology 479-480:131–145.

[70] Monajemi M, Woodworth CF, Benkaroun J, Grant M, Larijani M. 2012. Emerging complexities of APOBEC3G action on immunity and viral fitness during HIV infection and treatment. Retrovirology 9:Article 35

[71] Grant M, Larijani M. 2017. Evasion of adaptive immunity by HIV through the action of host APOBEC3G/F enzymes. AIDS Research and Therapy 14:Article 44

[72] Jarmuz A, Chester A, Bayliss J, Gisbourne J, Dunham I, Scott J, Navaratnam N. 2002. An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. Genomics 79:285–296.

[73] Conticello SG. 2008. The AID/APOBEC family of nucleic acid mutators. Genome Biology 9:Article 229

[74] Prohaska KM, Bennett RP, Salter JD, Smith HC. 2014. The multifaceted roles of RNA binding in APOBEC cytidine deaminase functions. Wiley Interdisciplinary Reviews: RNA 5:493–508

[75] Betts L, Xiang S, Short SA, Wolfenden R, Carter Jr CW. 1994. Cytidine deaminase. The 2.3 A crystal structure of an enzyme: transition-state analog complex. Journal of Molecular Biology 235:635–656.

[76] I.B. Rogozin, et al. APOBEC4, a new member of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases predicted by computational analysis. Cell Cycle, 4 (2005), pp. 1281-1285

[77] S.G. Conticello, et al. Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. Mol. Biol. Evol., 22 (2005), pp. 367-377

[78] Lorenzo, J. P. et al. APOBEC2 is a transcriptional repressor required for proper myoblast differentiation. Preprint at bioRxiv (2021)

[79] Marino, D. et al. APOBEC4 enhances the replication of HIV-1. PLoS ONE 11, e0155422 (2016).

[80] Shi, M. et al. Characterization and functional analysis of chicken APOBEC4. Dev. Comp. Immunol. 106, 103631 (2020).

[81] Melcher, T., Maas, S., Herb, A., Sprengel, R., Higuchi, M., and Seeburg, P.H. (1996). RED2, a brain specific member of the RNA-specific adenosine deaminase family. J. Biol. Chem. 271(50), 31795-31798.

[82] Schumacher, J. M., Lee, L., Edelhoff, S., and Braun, R. E. (1995). Distribution of TENR, an RNA-binding protein, in a lattice-like network within the spermatid nucleus in the mouse. Biol. Reprod. 52(6), 1274-1283.

[83] Kumar M, Carmichael GC: Nuclear antisense RNA induces extensive adenosine modifications and nuclear retention of target transcripts. Proc Natl Acad Sci USA 1997, 94:3542-3547.

[84] Hogg M, Paro S, Keegan LP, O'Connell MA. RNA Editing by Mammalian ADARs. 1st ed. Elsevier Inc., 2011 doi:10.1016/B978-0-12-380860-8.00003-3.

[85] Patterson JB, Samuel CE. Expression and regulation by interferon of a double-stranded RNA-specific adenosine deaminase from human cells: evidence for two forms of the deaminase. Mol Cell Biol 1995; 15: 5376–5388.

[86] Antje Keppler, Susanne Gendreizig, Thomas Gronemeyer, Horst Pick, Horst Vogel, and Kai Johnsson, A general method for the covalent labeling of fusion proteins with small molecules in vivo. Published online 9 December 2002; doi:10.1038/nbt765

[87] Thorsten Stafforst, Marius F. Schneider, An RNA–Deaminase Conjugate Selectively Repairs Point Mutations. Angew. Chem. Int. Ed. 2012, 51, 11166–11169, DOI: 10.1002/anie.201206489

[88] Paul Vogel, Marius F. Schneider, Jacqueline Wettengel, and Thorsten Stafforst, Improving Site-Directed RNA Editing In Vitro and in Cell Culture by Chemical Modification of the Guide RNA. Angew. Chem. Int. Ed. 2014, 53, 6267–6271. DOI: 10.1002/anie.201402634.

[89] Paul Vogel, Matin Moschref, Qin Li, Tobias Merkle, Karthika D. Selvasaravanan, Jin Billy Li & Thorsten Stafforst, Efficient and precise editing of endogenous

transcripts with SNAP-tagged ADARs. Nature Methods volume 15, pages 535–538 (2018).

[90] Tan, R. and Frankel, A.D. (1995) Structural variety of arginine-rich RNA-binding peptides. Proc. Natl. Acad. Sci. U.S.A. 92, 5282–5286.

[91] Friedman, D. and Gottesman, M. (1983) Lytic mode of lambda development, In Lambda II (Hendrix, R.W., Roberts, J.W., Stahl, F.W., and Weisberg, R.A., eds). Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

[92] Legault, P., Li, J., Mogridge, J., Kay, L.E. and Greenblatt, J. NMR structure of the bacteriophage lambda N peptide/boxB RNA complex: recognition of a GNRA fold by an arginine-rich motif, Cell, 1998, vol. 93 pp. 289–299.

[93] Maria Fernanda Montiel-Gonzalez, Isabel Vallecillo-Viejo, Guillermo A. Yudowski, and Joshua J. C. Rosenthal, Correction of mutations within the cystic fibrosis transmembrane conductance regulator by site-directed RNA editing. PNAS, November 5, 2013, vol. 110, no. 45, 18285–18290.

[94] M.F. Montiel-González, I.C. Vallecillo-Viejo, J.J. Rosenthal. An efficient system for selectively altering genetic information within mRNAs, Nucleic Acids Res. 44 (21) (2016).

[95] M. O'Connell, Molecular mechanisms of RNA-targeting by Cas13-containing typeVI CRISPR-cas systems, J. Mol. Biol. (2018).

[96] S. Shmakov, A. Smargon, D. Scott, D. Cox, N. Pyzocha, W. Yan, et al. Diversity and evolution of class 2 CRISPR–Cas systems. Nat. Rev. Microbiol., 15 (2017), pp. 169-182

[97] W.X. Yan, S. Chong, H. Zhang, K.S. Makarova, E.V. Koonin, D.R. Cheng, et al. Cas13d is a compact RNA-targeting type VI CRISPR effector positively modulated by a WYL-domain-containing accessory protein. Mol. Cell, 70 (2) (2018), pp. 327-339.e5

[98] S. Konermann, P. Lotfy, N.J. Brideau, J. Oki, M.N. Shokhirev, P.D. Hsu. Transcriptome engineering with RNA-targeting type VI-D CRISPR effectors. Cell, 173 (2018). (665-76 e14)

[99] D.B.T. Cox, J.S. Gootenberg, O.O. Abudayyeh, B. Franklin, M.J. Kellner, J. Joung, F. Zhang, RNA editing with CRISPR-Cas13, Science 358 (6366) (2017) 1019–1027.

[100] Xiaoqiang Wang, Phillip D. Zamore, Traci M.Tanaka Hall, Crystal Structure of a Pumilio Homology Domain, Molecular Cell, Volume 7, Issue 4, 2001, Pages 855-865,

[101] Cheong, C.-G. & Hall, T. M. T. Engineering RNA sequence specificity of Pumilio repeats. Proc. Natl Acad. Sci. USA 103, 13635–13639 (2006).

[102] Dong S, Wang Y, Cassidy-Amstutz C, Lu G, Bigler R, Jezyk MR, Li C, Hall TM, Wang Z. Specific and modular binding code for cytosine recognition in Pumilio/FBF (PUF) RNA-binding domains. J Biol Chem. 2011 Jul 29;286(30):26732-42.

[103] Han W, Huang W, Wei T, Ye Y, Mao M, Wang Z. Programmable RNA base editing with a single gRNA-free enzyme. Nucleic Acids Res. 2022 Aug 27;50(16):9580–95.

[104] Abudayyeh OO, Gootenberg JS, Franklin B, Koob J, Kellner MJ, Ladha A, Joung J, Kirchgatterer P, Cox DBT, Zhang F. A cytosine deaminase for programmable single-base RNA editing. Science. 2019 Jul 26;365(6451):382-386.

[105] Khosravi HM, Jantsch MF. Site-directed RNA editing: recent advances and open challenges. RNA Biol. 2021 Oct 15;18(sup1):41-50.

[106] Huang X, Lv J, Li Y, Mao S, Li Z, Jing Z, Sun Y, Zhang X, Shen S, Wang X, Di M, Ge J, Huang X, Zuo E, Chi T. Programmable C-to-U RNA editing using the human APOBEC3A deaminase. EMBO J. 2021 May 3;40(9):e108209. doi: 10.15252/embj.2021108209. Erratum for: EMBO J. 2020 Nov 16;39(22):e104741.

[107] Tang G, Xie B, Hong X, Qin H, Wang J, Huang H, Hao P, Li X. Creating RNA Specific C-to-U Editase from APOBEC3A by Separation of Its Activities on DNA and RNA Substrates. ACS Synth Biol. 2021 May 21;10(5):1106-1115.

[108] Cheng-Ting Chien, Paul L. Bartel, Rolf Sternglanz, and Stanley Fields, The two-hybrid system: A method to identify and clone genes for proteins that interact with a protein of interest. Proc. Nati. Acad. Sci. USA Vol. 88, pp. 9578-9582, November 1991 Biochemistry

[109] D.J. SenGupta, B. Zhang, B. Kraemer, P. Pochart, S. Fields, M. Wickens, Proc. Natl. Acad. A three-hybrid system to detect RNA-protein interactions in vivo. Sci. USA 93 (1996) 8496–8501.

[110] Brad Hook, David Bernstein, Beilin Zhang, and Marvin Wickens, RNA–protein interactions in the yeast three-hybrid system: Affinity, sensitivity, and enhanced library screening. RNA (2005), 11:227–233.

[111] Cecile Keryer-Bibens, Carine Barreau and H. Beverley Osborne, Tethering of proteins to RNAs by bacteriophage proteins. Biol. Cell (2008) 100, 125–138

[112] Ni, C.Z., Syed, R., Kodandapani, R., Wickersham, J., Peabody, D.S. and Ely, K.R. (1995) Crystal structure of the MS2 coat protein dimer: implications for RNA binding and virus assembly. Structure 3, 255–263

Chapter2 A-to-I RNA editing by using ADAR1 artificial deaminase system with MS2 12 X stem-loop

2.1 Introduction

The programmable nucleases for genome editing, such as CRISPR-Cas9, TALENs, RGENs, and ZFNs, hold promise for use in gene therapy [1]. However, genome editing has certain drawbacks, as changes made outside of target regions may be irreversible and can alter the genome's sequence [2]. In contrast, RNA editing results in mismatches that are not permanent and don't alter the genome's sequence. Therefore, RNA editing may be safer than genome editing for medical applications [3].

As a type of post-transcriptional modification, RNA editing enables the generation of multiple protein isoforms from a single gene [4]. The conversion from adenosine (A) to inosine (I) is one kind of the site-directed RNA editing. During translation, inosine (I) is erroneously recognized as guanosine (G). Thus, the replacement of A with G residues may led to restore amino acid codes, splicing elements, miRNAs, or miRNA-binding [6]. Besides, the AID-APOBEC enzyme family was used to accomplish the C-to-U conversion. Many 3'UTR targets of APOBEC1 have been identified before [7]. RNA editing can be used to correct specific mutations without permanently altering the genome's sequence, making it a potentially advantageous approach for medical applications.

A novel technique called "site-directed RNA editing" has been discovered, which allows for the recoding of single A residues at specific locations in any desired transcript

by using engineered deaminases in combination with short guide RNAs [8-11]. The targeting and specificity of the process can be programmed using guide RNAs through the use of Watson-Crick base pairing principles [12, 13].

A range of guide RNA types have recently been used to increase the specificity of the enzyme's targeting. In earlier studies, the deaminase domain of human ADAR2 was fused with the lambda-N peptide (lambda-N plus ADAR2) to correct CFTR (cystic fibrosis transmembrane conductance regulator) mutations [14]. The chemically modified guide RNAs have all been utilized to fix point mutations in annelids and human cell lines [13]. Previously, we achieved the chemical RNA editing by using oligodeoxynucleotides (ODNs) that containing 5'-carboxyvinyl-2'-deoxyuridine [(CV)U] [15, 16]. More recently, a genetically encoded guide RNA was utilized to correct point mutations related to a neuronal disease phenotype [17].

In the present study, I attempted to replace the 6X MS2 stem-loop with a 12X MS2 stem-loop in order to achieve higher editing efficiency (**Fig 1**). Consider that the MS2 system is more commonly used than lambda N/B box systems for attaching proteins to RNAs [18], this approach may become a useful tool to label RNAs with GFP *in vivo* [19].
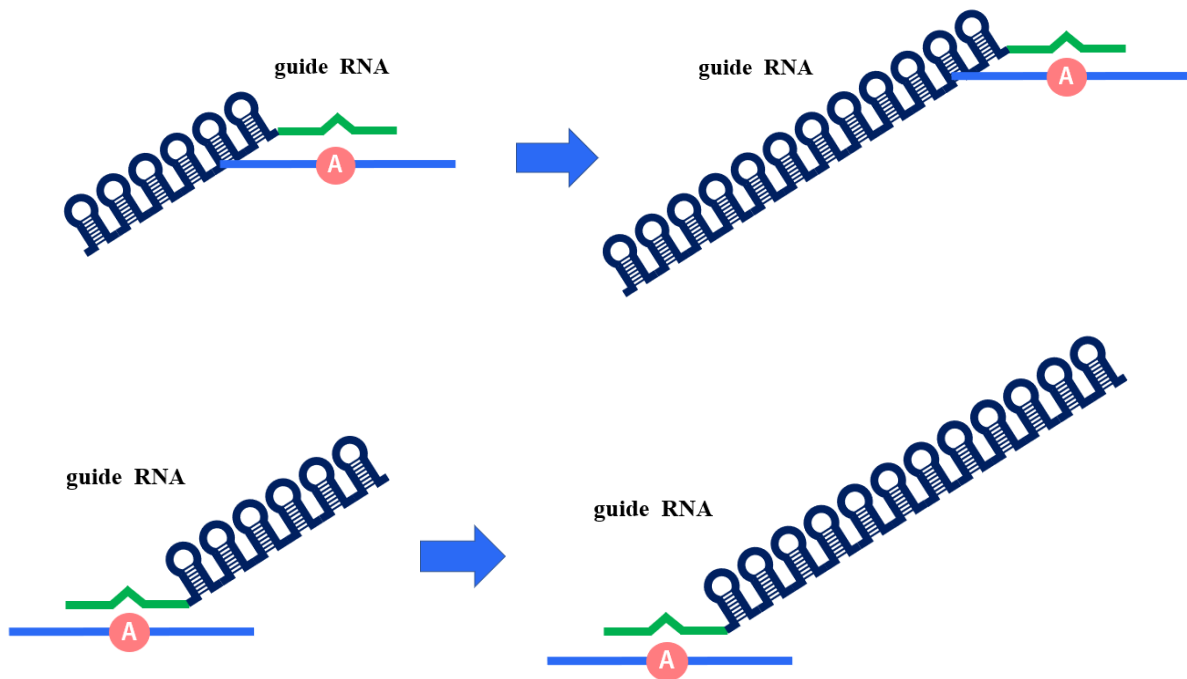
Fig. 1: Schematic feature of the 12 X MS2 stem-loop guide RNA. In this study, we replaced the 6 X MS2 stem-loop RNA into 12 X MS2 stem-loop. We also designed the upstream and downstream guide RNA for further investigation.

2.2 Materials and Methods

### 2.2.1 Plasmid construct preparation

In this study, the key objective is to generate the PCS2+ 12 X MS2 stem-loop RNA construct (**Fig 1**), which enables the precise targeting and introduce the enzyme to the specific site. Previously, my colleague fused the MS2 protein upstream of the ADAR1 deaminase domain to create a new plasmid (PCS2+MT+MS2HB+ADAR1) by fused the MS2 protein on the upstream of the ADAR1 deaminase domain (**Fig 2**). Additionally,

they fused the guide RNA upstream of the MS2 RNA to obtain the other vital plasmid

(PCS2+guide RNA+ 6 X MS2 stem-loop RNA) (**Fig 2**).
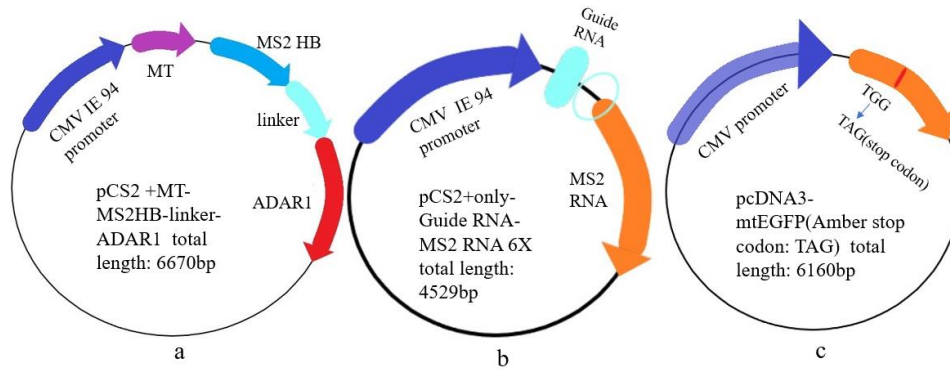
Fig 2: Plasmid constructs used in my study (Acknowledge: I thank Dr. Thoufic for the

providing of the wild plasmids)

We used EcoR I (TaKaRa, catalog number: 1040A) and Xba I (TaKaRa, catalog

number: 1093A) to digest PCS2+ (Addgene) as vector. Next, we used EcoR I (BioLabs,

catalog number: R0101) and Nhe I (BioLabs, catalog number: R0131) to digest pSL-

MS2-12X (Addgene) for obtaining the insertion fragment. After restriction digestion, the

products were run by 1% agarose gel, then cut the gel and extracted the target fragment

by using NucleoSpin gel and PCR clean-up kit (TaKaRa). We performed the

dephosphorylation by using Alkaline Phosphatase (TaKaRa, E. coli C75, catalog number:

2120A) on the vector. After obtaining the vector and insert, we ligated them by T4 DNA

ligase (BioLabs, Catalog number: M0202). (Scheme1)
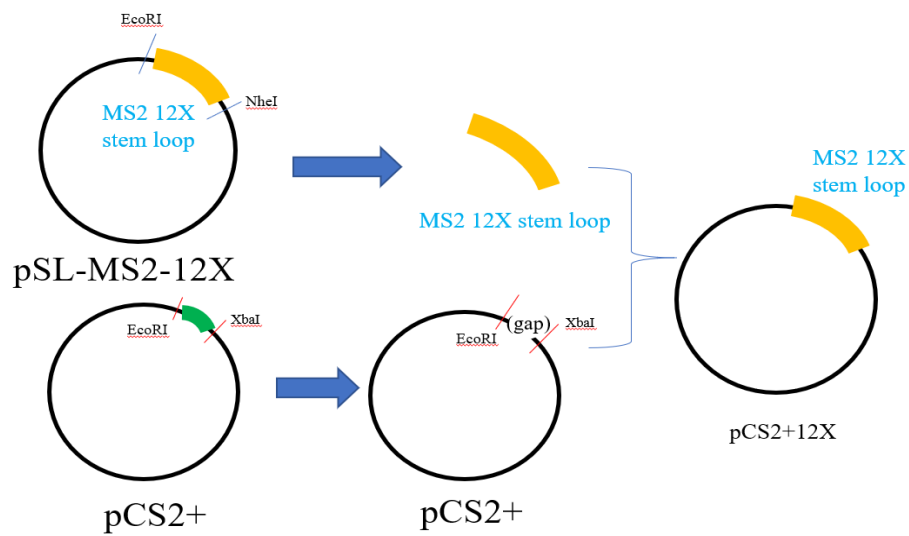
Scheme1: Preparation of PCS2+12 X MS2 stem-loop RNA

The reason for choosing Nhe I and Xba I is that they are isocaudomers, which means, even recognize different DNA sequences, both of the restriction enzymes generate the same sticky ends after cleavage and easy ligation (**Fig. 3**). However, the disadvantage of using isocaudomers is that the resulting asymmetric sequence after ligation is usually no longer recognized by the same restriction enzyme.
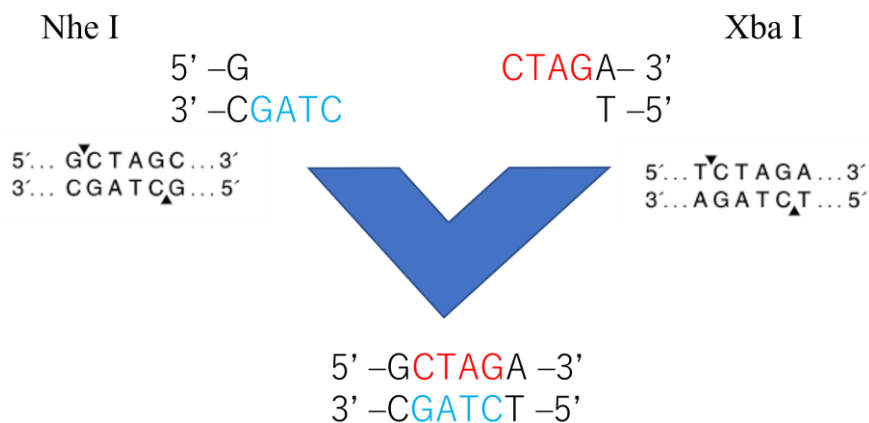
Fig 3: Scheme of ligation. The enzymes Nhe I and Xba I are isocaudomers. They can generate the same sequence tail on the target site.

After ligation, we transformed the ligation mixture into DH5α competent cells. 16 hours later, picked up the clones and cultured in liquid LB broth for 16 hours. Next, we extracted the plasmid by QIAGEN Plasmid Midi Kit (catalog number: 12143). The concentration confirmation was performing by using NanoDrop® ND-1000 (Thermo Scientific). Then, we used EcoRV (TaKaRa, catalog number: 1042A) for restriction digestion for construct confirmation. We used specific primer to perform sequencing confirmation (Applied Biosystems 3130xl Genetic Analyzer, USA; specific primer: MS2 RNA fwd---GATTACGAATTCGAATGGCCATG).

2.2.2 guide RNA insertion

After the preparation of PCS2+MS2 RNA 12X, we inserted the complementary sequence into the vector. Referred to the previous studies, I selected the guide 21 and guide 23 as complementary sequence in my study. I designed the upper strand as forward primer (Fwd) and bottom strand as reverse primer (Rev). The sequences were in the **Table 1.** (Eurofins)

Guide21 Fwd	ATC GAGGGGGGGGCCAGGGCACGGG GAT
Guide21 Rev	ATC CCCGTGCCCTGGCCCCCCCTC GAT
Guide23 Fwd	ATC CGAGGGGGGGGCCACCGCACGGGC GAT
Guide23 Rev	ATC GCCCGTGCGGTGGCCCCCCCTCG GAT

Table1: Guide RNA was designed with insertion either to the upstream or downstream of PCS2+12 X MS2 stem-loop RNA. Green mark and under line is guide RNA sequence. Red color

indicates CCA and TGG which are targeting to mutated codon. Yellow color indicates C and G

mismatch which are adopted to avoid off-target editing.

After annealing thermal cycle (**Table 2**), the confirmation was performed by using

polyacrylamide gel electrophoresis.

Temperature	94°C	60°C	50°C	25°C	4°C
Time	2min	2min	2min	2min	Keep

Table2: Annealing thermal cycle

Because there is not phosphate group on oligo-synthesized DNA at the 5'-OH end,

we performed the phosphorylation on the annealed fragment. T4 Polynucleotide Kinase

(TaKaRa, catalog number: 2021S) was used for phosphorylation. The thermal cycle was

shown in **Table 3**.

Temperature	37°C	37°C	70°C	4°C
Time	30sec	45min	5min	Keep

Table3: Phosphorylation thermal cycle

After phosphorylation, we performed the ethanol precipitation for ligation reaction.

Next, we designed upstream guide and downstream guide for the PCS2+12 X MS2

stem-loop RNA. For the upstream guide, we used EcoR I (TaKaRa, **Fig 4**) and Nco I

(TaKaRa, **Fig 4**) to digest the vector. For performing the blunt end ligation, we used the

Klenow Fragment (TaKaRa, Large Fragment E. coli DNA Polymerase I, catalog number:

2140A) for the blunt end treatment on vector.

For the downstream guide, we used EcoR V (TaKaRa, catalog number: 1042A, **Fig 4**) to digest the vector.
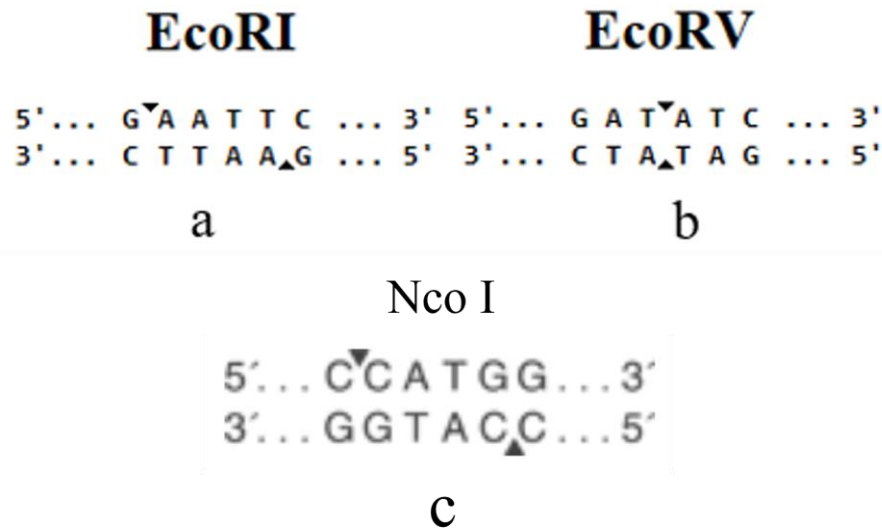
Fig 4: a) EcoR I restriction site; b) EcoR V restriction site. c) Nco I restriction site.

After restriction digestion and blunt end treatment, both vectors performed the dephosphorylation by using Alkaline Phosphatase (TaKaRa, E. coli C75, catalog number: 2120A) and phenol chloroform ethanol precipitation.

The ligation of the vector and insert was performed by using T4 DNA ligase (BioLabs, Catalog number: M0202). (**Scheme2**)
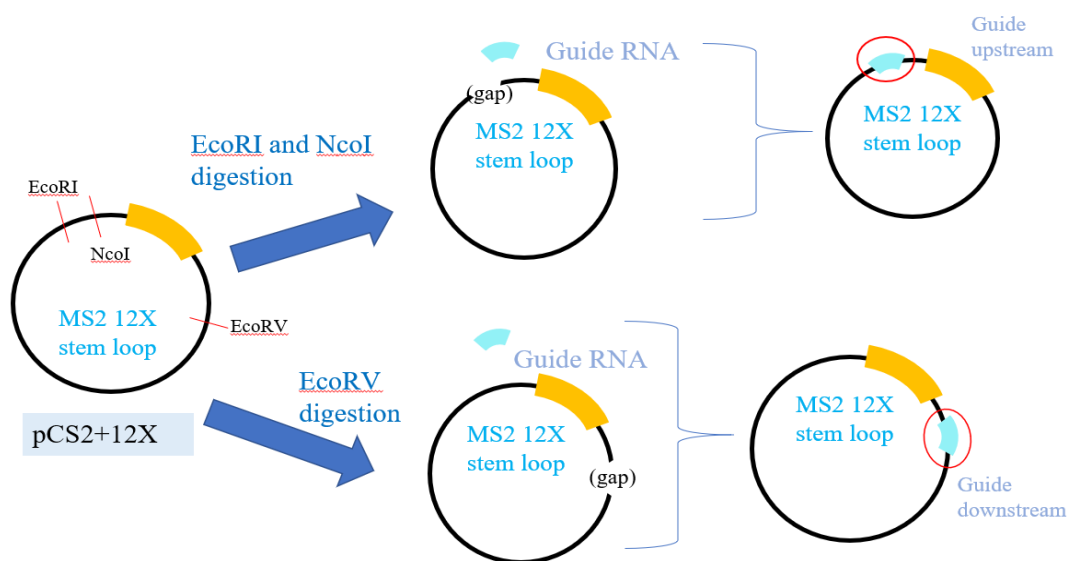
Scheme2: Protocol of upstream and downstream guide RNA

After ligation, we transformed the 2μL of the ligation mix into DH5α competent cells. 16 hours later, we picked up the clones and cultured in the liquid LB broth for 16 hours. Next, we extracted the plasmid by using QIAGEN Plasmid Midi Kit (catalog number: 12143). We confirmed the concentration by using NanoDrop® ND-1000 (Thermo Scientific). For upstream guide RNA, we used specific primer (**Table 4**) in PCR to confirm the ligation result. For downstream guide RNA confirmation, we used Hpa I and Bgl II double digestion (**Fig. 5**). The results were confirmed by polyacrylamide gel electrophoresis and sequencing.

PCS2 12X fwd	AGCAAGCTTGATTTAGGTGACA
PCS2 12X rev	CATATATAGGGCCCGGGTTATAA

**Table4: Specific primer for upstream guide RNA confirmation**
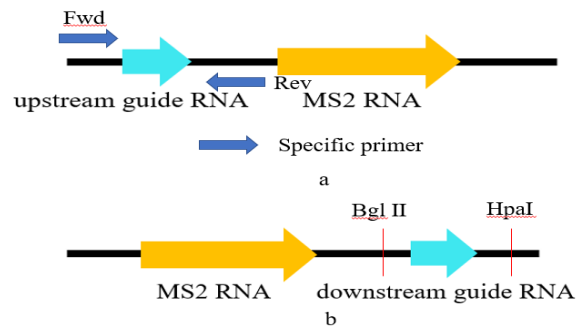
Fig 5: a) The location about specific primer; b) Hpa I and Bgl II restriction on the plasmid construct.

One disadvantage of blunt end ligation is that the direction of the insert may be reverse. So, we performed the sequencing for confirming the direction of the guide RNA on 3130xl Genetic Analyzer (Applied Biosystems, USA).

## 2.2.3 Construction of target and reporter substrate

In this study, mutated EGFP was used as a target of the RNA editing system (**Fig 6**). This construct was used in previous study. Mutations were introduced at the 58th amino acid position (TGG to TAG, encoding amber stop codon) by using a site-directed mutagenesis kit.
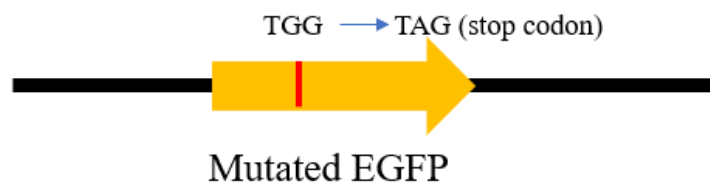
Fig 6: Mutated EGFP (Amber stop codon)

2.2.4 Cell culture and transfection

Approximately, 4×10^5 cells/well were seeded in the 12-well culture plates (FALCON, USA), grown for 24h to 80-90% confluence, and then subjected to transfection. The medium of D-MEM with high glucose (nacalai tesque, catalog number: 09891-25) supplemented with 10% fetal bovine serum (Gibco, catalog number: 26140079) was used for cell culture. The polyethylenimine (PEI) was utilized for transfections (**Scheme 3**). (Acknowledge: I thank Dr. Thoufic for providing of the wild plasmids)
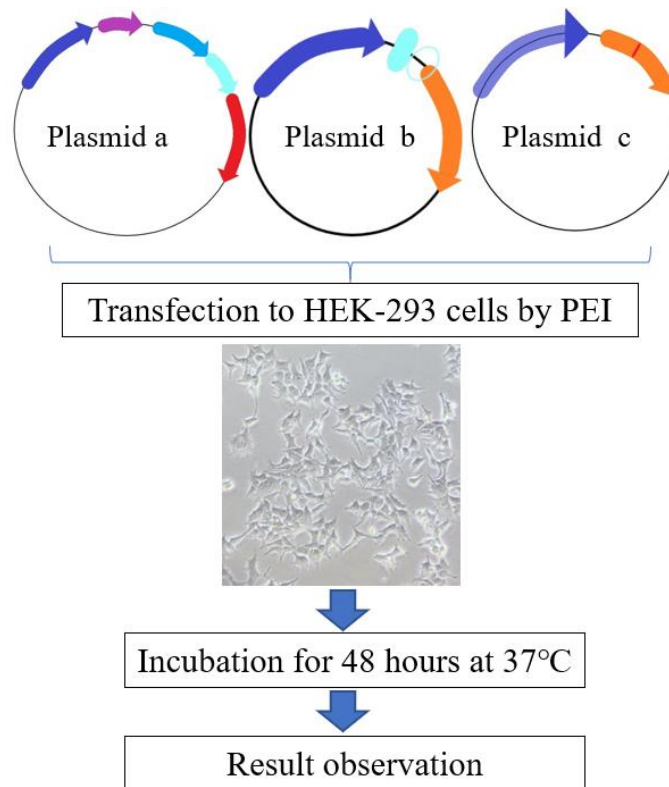
Scheme3: Transfection and observation

2.2.5 Cell observation

48 hours later, fluorescent signals were observed by using a fluorescence microscopy (Keyence Biozero-800) as per standard conditions.

2.2.6 RNA extraction and cDNA synthesis

Trizol reagent (Invitrogen) was used to extract total RNA from the cells. Firstly, the culture medium was aspirated, and 200µl of Trizol reagent was added to each well. The cells were further lysed by directly pipetting up and down several times in the culture dish. After that, the lysate was transferred to 1.5 ml tubes. Subsequently, 200 µl of chloroform was added. All the sample was shaken hard by hand for 15 seconds followed by 3-minute incubation at room temperature. All the sample was centrifuged at 12000 × g for 15 minutes at 4℃. The aqueous supernatant was transferred to a new tube and mixed with 500 µl of isopropanol and incubated at room temperature for 10min. For precipitation of RNA, followed by centrifugation at 12000 × g for 10 minutes at 4℃ and the supernatant was removed. The RNA precipitation was washed with 0.5 ml of 75% ethanol, and then centrifuged at 7500 × g for 5 minutes at 4℃. Finally, the supernatant was removed, and the RNA precipitation was stored in room temperature for air-dry resuspended in DEPC-treated water (Invitrogen, catalog number: R0603).

Next, for removing all the interference from DNA, RQ1 RNase-Free DNase (Promega, catalog number: M6101) was used. After DNase treatment, about 1µg of RNA was utilized for cDNA synthesis by using Oligo(dT) 15nt primer (Promega, catalog number: C110A) and the Superscript IV reverse transcriptase kit (Invitrogen, catalog number: 18090010) for reverse transcription.

2.2.7 Confirmation of restoration

After obtaining the cDNA, we performed the PCR reaction by using specific primer. PCR was performed by using *Ex Taq* (TaKaRa, catalog number: RR001A). Every experiment was repeated least three times. Thermal cycling conditions were as **Table 5**.

Step	Temperature	Time
Initial Denaturation	95°C	3 minutes
Denaturation	95°C	30 seconds
Annealing	60°C	30 seconds (total cycle: 30)
Extension	72°C	2 minutes
Final Extension	72°C	2 minutes
Soak	4°C	Indefinite

Table5: PCR thermal cycle

Specific primers were designed (**Table 6**) and ordered from Eurofins. The total length of the PCR product is 324bp.

324 EGFP fwd	AAGTTCAGCGTGTCCGGC
324 EGFP rev	GTCCTCCTTGAAGTCGATGC

Table6: Specific primers for PCR

The PCR products were ethanol precipitated and quantitated by NanoDrop® ND-1000 (Thermo Scientific). Then, electrophoresis by 6% polyacrylamide gel and stained with SYBR green (Lonza) for confirming the target band.

PCR products can be digested with a restriction enzyme that will generate different length of fragments (the edited and unedited DNA fragments). The mutated EGFP sequence has been confirmed on NEBcutter V2.0 (BioLabs), and there is only one restriction site by Bfa I on the PCR products. After confirmation, equal amounts of PCR products were digested by Bfa I at 37℃ for 2 h. Bfa I can digest the 324bp PCR product into two fragments of 231 and 93 bp. For the mutated allele of EGFP, Bfa I can digest (5'···CCCTAGCC···3') due to the presence of the amber stop codon at the 58th amino acid that different from the edited one---5'···CCCTI(G)GCC···3' (**Fig 7**). The restriction digestion products were also run in 6% polyacrylamide gel and imaged by using LAS-3000 Imager (Fujifilm).
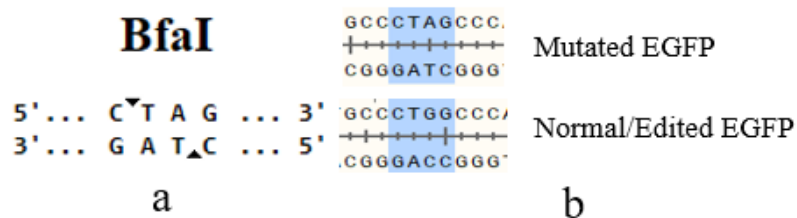
Fig 7: a) Bfa I restriction site sequence; b) Difference between mutated EGFP and normal/edited EGFP.

2.2.8 Sequencing of PCR products for editing efficiency observation

The Genetic Analyzer of 3130xl was used for sequencing of amplicons. After running the PCR products on 1% agarose gels and staining them with ethidium bromide

gel stain solution, the bands were cut and purified with the NucleoSpin gel and PCR

clean-up kit from TaKaRa. The concentration of the purified products was determined

with a NanoDrop® ND-1000 (Thermo Scientific) spectrophotometer. Finally,

sequencing was performed using the kit of BigDye® Terminator v3.1 Cycle (catalog

number: 4337455) with the thermal cycling conditions listed in **Table 7**.

Temperature	Time
96°C	1min
96°C	10sec
50°C	5sec
60°C	4min
4°C	∞

Table7: Sequence Thermal cycle

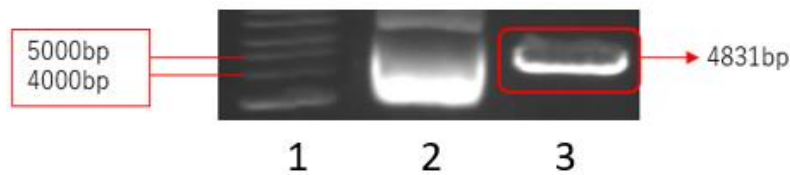# 2.3 Results

2.3.1 Result of the PCS2+MS2 RNA 12X construct plasmid preparation

Fig 8: Agarose gel electrophoresis result of pSL-MS2-12X was digested by EcoR I and Nhe I (Marker: 100bp DNA Laddar); Picture was taken by Vilber lourmat transilluminator.

As shown in **Fig 8**, 12 X MS2 stem loop fragment had been cut from pSL-MS2-12X plasmid and electrophoresis by 1% agarose gel. The length of the fragment was 754bp.

After extracting by QIAGEN Plasmid Midi Kit, used EcoR V (TaKaRa, catalog number: 1042A) for restriction analysis. The reason that EcoR V restriction site was only exists on MS2 RNA 12 X fragment. (**Fig 9**)

Fig 9: Agarose gel electrophoresis results of PCS2+MS2 RNA 12X. Lane1: 1kb DNA ladder; Lane2: PCS2+MS2 RNA 12X without restriction digestion; Lane3: PCS2+MS2 RNA 12X restriction digestion by EcoR V. Picture was taken by Vilber lourmat transilluminator.

Sequencing result was analyzed by Applied Biosystems 3130xl Genetic Analyzer (USA) by using specific primer. (bold, underlined italics, MS2 stem loop; underlined, PCS2 vector backbone)

Analyzed result:

GGTTCATCAGTATTATAACCCGGGCCTATATATGGATCCTAAGGTACCTAA

TTGCCTAGAAA***ACATGAGGATCACCCATGT***CTGCAGGTCGACTCTAGAAA***AC***

*__ATGAGGATCACCCATGT__*CTGCAGTATTCCCGGGTTCATTAGATCCTAAGGTAC

CTAATTGCCTAGAAA*__ACATGAGGATCACCCATGT__*CTGCAGGTCGACTCTAGAA

A*__ACATGAGGATCACCCATGT__*CTGCAGTATTCCCGGGTTCATTAGATCCTAAGG

TACCTAATTGCCTAGAAA*__ACATGAGGATCACCCATGT__*CTGCAGGTCGACTCCA

GAAA*__ACATGAGGATCACCCATGT__*CTGCAGTATTCCCGGGTTCATTAGATCCTA

AGGTACCTAATTGCCTAGAAA*__ACATGAGGATCACCCATGT__*CTGCAGGTCGACT

CTAGAAA*__ACATGAGGATCACCCATGT__*CTGCAGTATTCCCGGGTTCATTAGATC

CTAAGGTACCTAATTGCCTAGAAA*__ACATGAGGATCACCCATGT__*CTGCAGGTCG

ACTCTAGAAA*__ACATGAGGATCACCCATGT__*CTGCAGTATTCCCGGGTTCATTAG

ATCCTAAGGTACCTAATTGCCTAGAAA*__ACATGAGGATCACCCATGT__*CTGCAGG

TCGACTCCAGAAA*__ACATGAGGATCACCCATGT__*CTGCAGTATTCCCGGGTTCAT

TAGATCTGCGCGCGATCGATATCAGCGCTTTAAATTTGCGCATGCTA<u>GAACTAT</u>

<u>AGTGAGTCGTATTACGTAGATCCAGA</u>CATGATAAGATACATTGATGAGTTTGG

ACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTGTG

ATGCTATTGCTTTAATTTGTAACCATTATAAGCTGCATAAAACAGTACACACAT

GCATTCATTTTTATGTTCAGGGTTCAGGGAAGGGGTGTGTGGGGAAGGGTTT

TTT

From the sequencing results, the MS2 RNA 12X was successfully inserted into the pCS2 plasmid.

2.3.2 Confirmation of upstream and downstream guide insertion

After obtaining the construct plasmid of PCS2+MS2 RNA 12X, we constructed the upstream and downstream guide RNA plasmid. For confirming the upstream guide RNA construction, I used the specific primer for PCR reaction. The PCR products were run by 6% polyacrylamide gel. Image was taken by LAS-3000 Imager (Fujifilm, **Fig 10**).
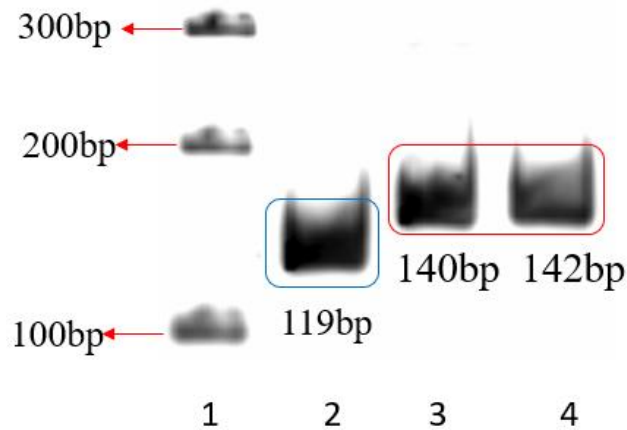
Fig 10: polyacrylamide gel electrophoresis result of upstream guide RNA (Lane1: 100bp DNA ladder, Lane2: no guide RNA insert vector; Lane3: guide RNA 21 insert; 4: guide RNA 23 insert)

Confirmation of downstream guide RNA was performed by Hpa I and Bgl II double digestion. Restriction digestion products were run on 6% polyacrylamide gel. Image was taken by LAS-3000 Imager (Fujifilm, **Fig 11**).
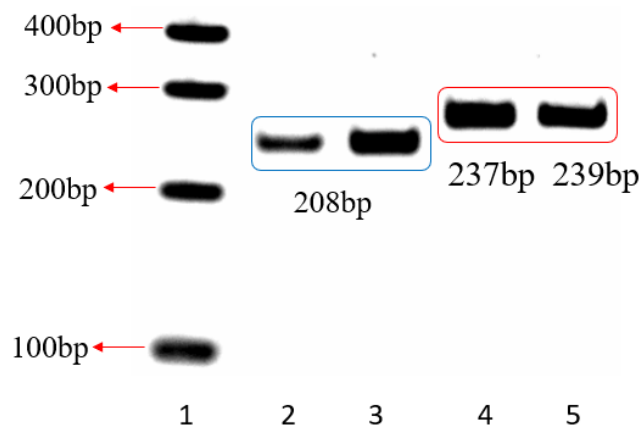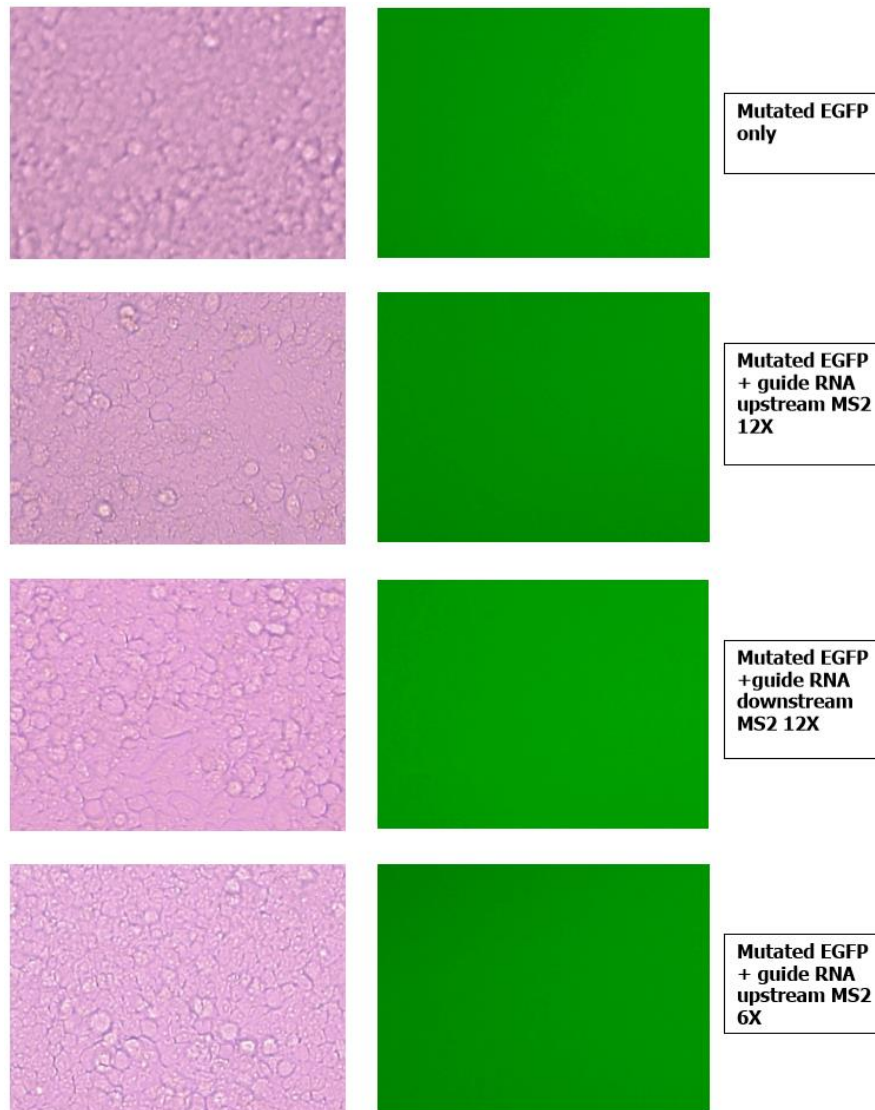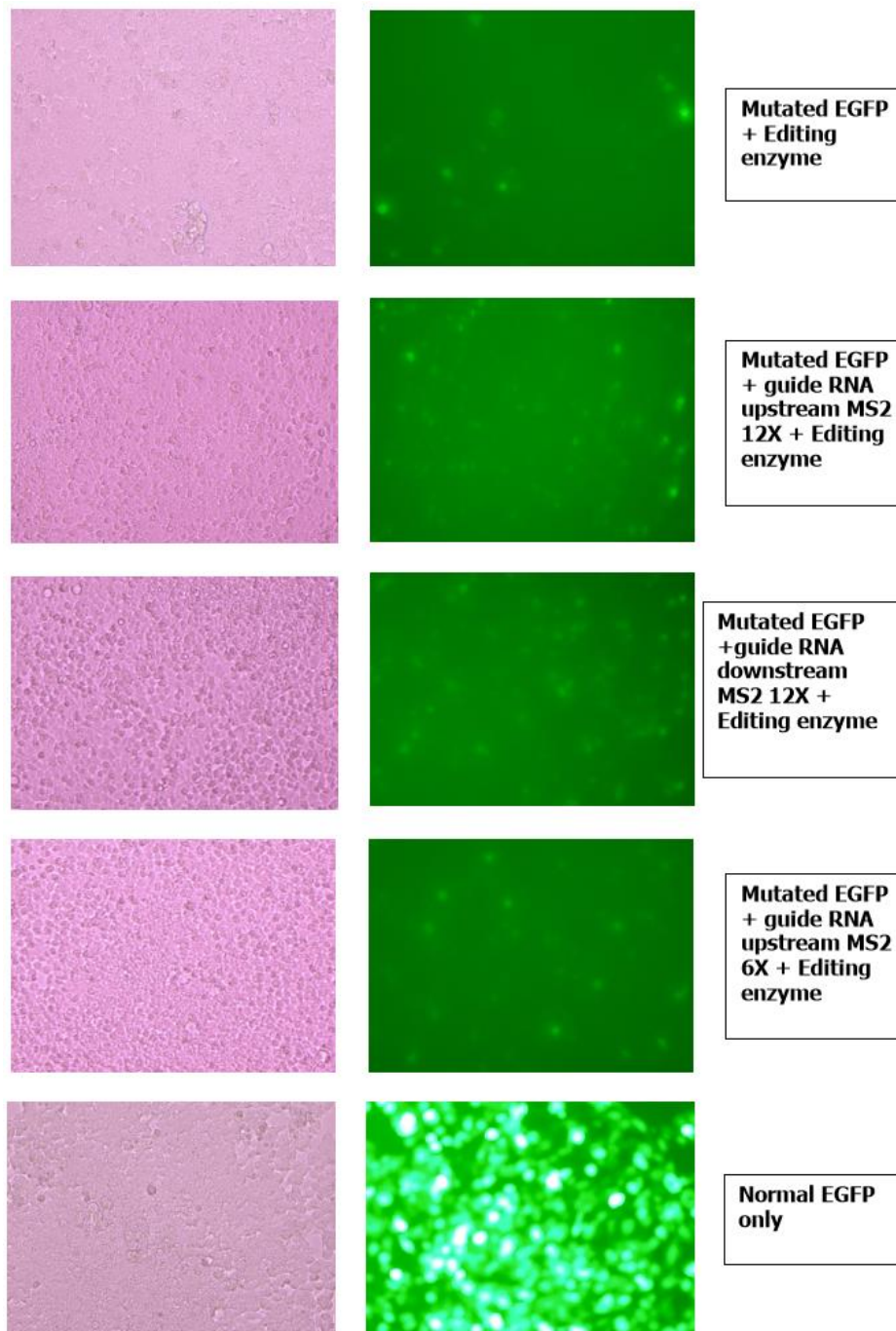
Fig 11: downstream guide RNA double digestion by Hpa I and Bgl II and PAGE result, Lane1: 100bp DNA ladder, Lane2 and 3: no guide RNA insert vector, Lane4: guide RNA 21 construct; Lane 5: guide RNA 23 construct.

From the electrophoresis results, I selected the sample in which the upstream and downstream guide RNA insertion is successful for sequencing to confirm the direction of the guide RNA. Then, the correct direction of the guide RNA constructs will be used for cell transfection experiment.

2.3.3 Transfection result

48 hours later, images were taken by Keyence Biozero-800 fluorescence microscope (**Fig 12**).

Mutated EGFP only

Mutated EGFP + guide RNA upstream MS2 12X

Mutated EGFP +guide RNA downstream MS2 12X

Mutated EGFP + guide RNA upstream MS2 6X

Fig 12: Cell images were taken by Keyence fluorescence microscope with different factors after transfection. Transfection factors we used are indicated on the right side of the images. The left side of images are phase contrast images, and the right side of images are fluorescence images. Each experiment was performed at least three times.

Fig. 12 shows that the green fluorescent signal could not be detected when either the mutated EGFP was transfected without editing enzyme. But a tiny signal of green fluorescent can still be observed when the guide RNA and the ADAR1 were transfected together.

We observed the intensity of fluorescent signals of 12 X MS2 stem-loop RNA and 6 X MS2 stem-loop RNA were almost at the same level. To further confirm the results, we performed RFLP (Restriction fragment length polymorphism) and cDNA sequencing.

2.3.4 Confirmation of genetic restoration

The aim in my study was confirming the efficiency about the genetic restoration. Before confirmation of the cDNA sequence, we performed RT-PCR. After that, RFLP analysis was conductedby using the *Bfa* I restriction enzyme. *Bfa* I can recognize and digest the 5'···CTAG···3' sequence (**Fig 7**). Thus, *Bfa* I cannot digest edited EGFP into two fragments (231 and 93 bp), but the not edited EGFP sequence will be digested into two fragments. (**Fig 13**)
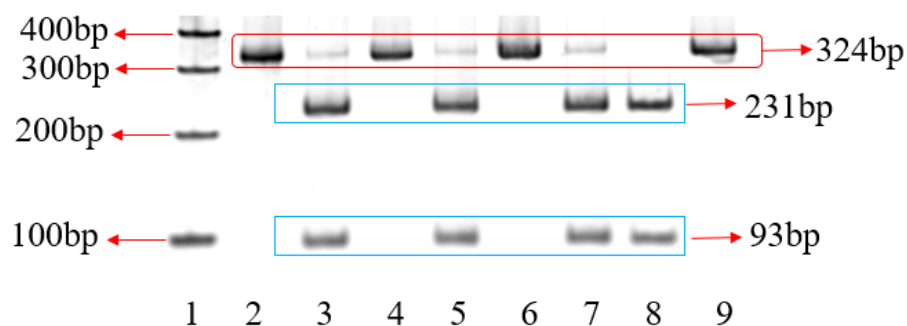

Fig 13: Restriction fragment length polymorphism (RFLP) analysis of RNA editing on mutated EGFP by PAGE. Lane1: 100bp DNA ladder; Lane2: MS2 RNA 12X- guide RNA

upstream RT-PCR product without restriction digestion; Lane3: MS2 RNA 12X- guide RNA

upstream RT-PCR product digested by Bfa I; Lane4: MS2 RNA 12X- guide RNA downstream

RT-PCR product without restriction digestion; Lane5: MS2 RNA 12X- guide RNA down-stream

RT-PCR product digested by Bfa I; Lane6: MS2 RNA 6X RT-PCR product without restriction

digestion; Lane7: MS2 RNA 6X RT-PCR product digested by Bfa I; Lane8: Mutated EGFP RT-

PCR product digested by Bfa I ; Lane9: Normal EGFP RT-PCR product digested by Bfa I.

From the results of the PAGE in Fig 12, we observed wild type EGFP   cannot be digested by restriction enzymes into two fragments, the mutated EGFP was completely digested into two fragments. From lane 3, 5 and 7, there were not only the 231 bp and 93 bp fragments, but also 324bp fragment that can't be digested by *Bfa* I.

2.3.5 Confirmation of RNA editing efficiency

After RFLP confirmation, we performed the DNA sequencing. For PCR products sequence, we used reverse primer (324 EGFP rev, **Table 6**) as sequencing primer. Then, we calculated the peak height ratio of C (cytosine) to T (Thymine) in the sequence result for obtaining the editing efficiency [21]. I performed 7 times PCR products direct sequencing of two transfected cells groups. The editing efficiency results had been showed in **Fig 14**. The editing efficiencies of guide RNA 21 and guide RNA 23 were almost at the same level. There was no significant difference in editing efficiencies observed between the 6X MS2 stem-loop RNA and 12X MS2 stem-loop RNA.
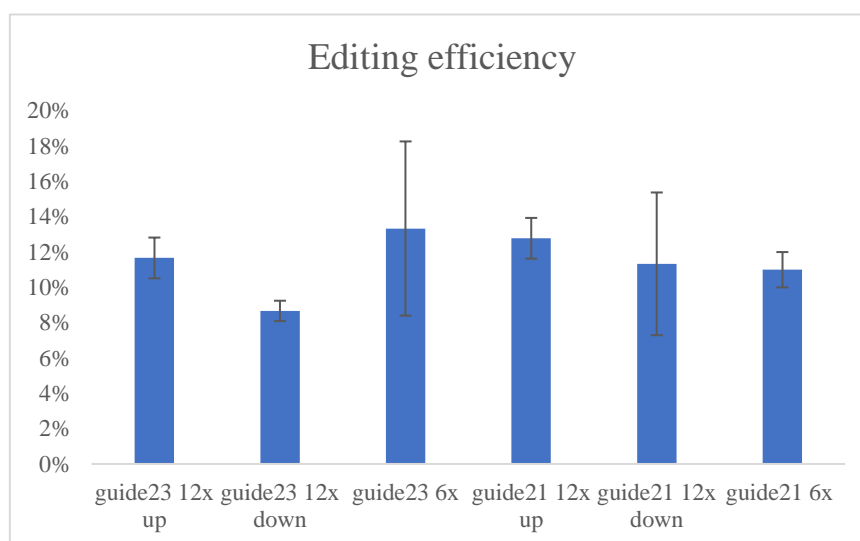
Fig 14: Editing efficiency results of all conditions.

In **Fig 14**, we observed that the editing efficiencies of each condition was not significant. There are many tolerance data in the sequencing results, and the error bar of the editing efficiencies were very large. We concluded the editing efficiency of upstream and downstream was at the same level in this study. And the editing efficiency of 12 X MS2 stem-loop RNA and 6 X MS2 stem-loop RNA was at the same level in this study.

2.5 Discussion

In the present study, we validated the possibility of the 12 X MS2 stem-loop RNA to increase the editing efficiency of the artificial RNA editing system. We also sought to investigate whether the position of guide RNA could affect editing efficiency or not.

When tethered proteins interact with RNA, protein binding sites on RNA can play an vital role. For instance, the fusion of two BoxB hairpins with the guide RNA of the

λN-BoxB system can enhance its editing efficiency [21]. Both the MS2 coat protein and the lambda-N protein are RNA-binding proteins from bacteriophages and have similarities in their RNA-protein interactions [13].

The Gorgoni group has ever found that by increasing the number of MS2-binding sites they could not improve the translational stimulation provided by the histone hairpin-binding protein [20]. However, for the RNA-binding protein Daz1, the translation of the reporter mRNA increased with the number of MS2 binding sites [22]. It is benefit to increase the number of binding sites on the RNA, if the concentration of tethered protein in the cell is limited. Nonetheless, studies have revealed that increasing the number of MS2-binding sites can lead to a significant decrease in the constant level of the reporter mRNA [18, 23]. The reasons for these obvious differences are unclear, but they may origin from to the intrinsic stability of the different reporter mRNAs used and the cell culture environment.

The results of the Stephanie group's study have shown that the introduction of PP7 stem-loops (PP7SL) or MS2 stem-loops (MS2SL) can influence the processing and subcellular localization of mRNA [24]. Therefore, we cannot disregard the effect of increasing the number of stem-loops on editing efficiency and the target mRNA in general.

Former studies have revealed that the editing efficiency of guide RNAs at the 3'-end (downstream guide) and 5'-end (upstream guide) of the stem-loop structure differs. Research results have shown that the editing efficiency of the 5'-guide is higher than that of the 3'-guide both in vivo and in vitro [25]. Moreover, the previous work of my colleague has also demonstrated that in the case of MS2 RNA 6X, the editing efficiency of the upstream guide RNA is higher [26].

Although my research data is unable to support the hypothesis that the editing efficiency of the upstream guide RNA is higher than that of the downstream guide RNA, I believe that in repeated experiments, the editing efficiency of the upstream guide RNA will be higher.

In my colleague's work, guide RNA 21 and 23 were confirmed to be more efficient. Specifically, the editing efficiency of guide RNA 21 was higher than that of guide RNA 23 [26]. However, in my experimental data, the editing efficiency of guide RNA 21 with MS2 RNA 12X is higher than that of guide RNA 23, but it is relatively low in the case of MS2 RNA 6X. This contradicts the data obtained by my colleague. One possible reason for this discrepancy is the distance between the MS2 stem-loop RNA and guide RNA, as shown in **Fig 14**. A long distance between the guide RNA and MS2 stem-loop RNA could potentially hinder ADAR1 from binding to the editing site.

guide RNA
upstream

12 X stem-loop

74nt length distance
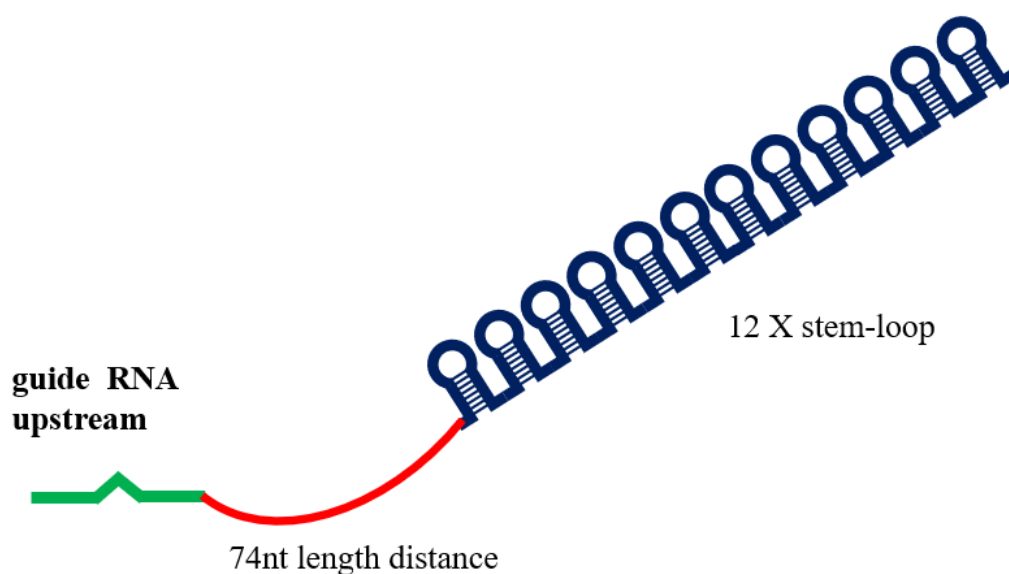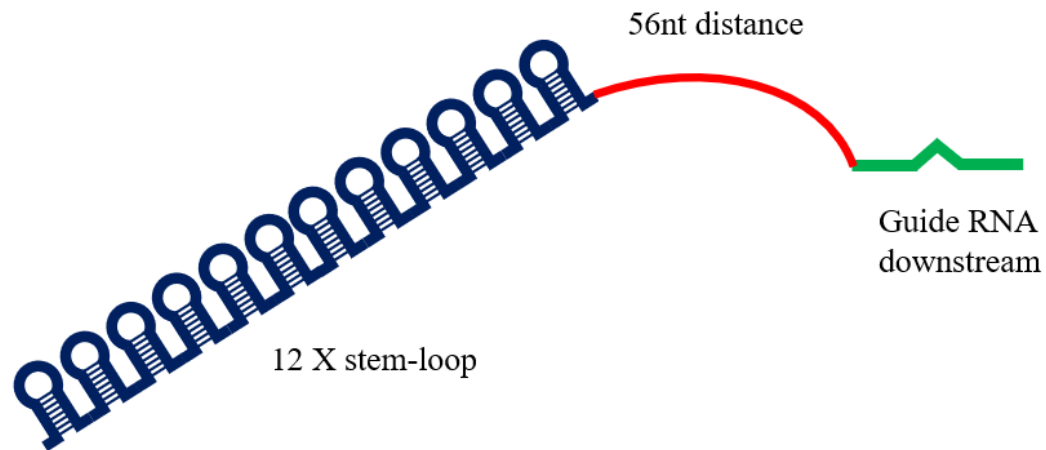
Fig 14: The distance between the guide RNA and MS2 stem-loop RNA in upstream or downstream guide RNA construct

In my transfection protocol, I used a total ratio of 1:1:1 for guide RNA, ADAR1, and the target EGFP plasmid, which resulted in equal proportions for all three components. My colleague's work also demonstrated that increasing the concentration of ADAR1 can improve the editing efficiency, but the editing efficiency decreases when concentration of guide RNA was increased [26]. This may be attributed to the effect of the stem-loop structure, as mentioned previously [24].

2.6 Conclusion

The results in this research suggest that the improvement in the artificial deaminase system was not significant. Specifically, MS2 RNA 12X can improve the editing efficiency of the system, but this improvement may also be influenced by the distance between MS2 stem-loop RNA and the guide RNA. Therefore, it is necessary to confirm whether MS2 RNA 12X or MS2 RNA 6X has a greater impact on the target mRNA, to avoid any unreasonable results due to excessive suppression of the target mRNA.

2.7 Reference

[1]   Kim H, Kim JS. A guide to genome engineering with programmable nucleases. Nat Rev Genet. 2014 May;15(5):321-34.

[2]   Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 2012; 337(6096): 816-821.

[3]   Leisegang M, Engels B, Schreiber K, Yew PY, Kiyotani K, Idel C et al. Eradication of large solid tumors by gene therapy with a T-cell receptor targeting a single cancer-specific point mutation. Clin Cancer Res 2016; 22(11): 2734-2743.

[4]   Maas S, Rich A. Changing genetic information through RNA editing. BioEssays 2000; 22(9): 790-802.

[5]   Vogel P., and Stafforst T. Site-directed RNA editing with antagomir deaminases-A tool to study protein and RNA function. Chem Med Chem, 9(9): 2021–2025, (2014).

[6]   Nishikura K. Functions and regulation of RNA editing by ADAR deaminases. Annu. Rev. Biochem, 79, 321–349, (2010).

[7]   Rosenberg BR, Hamilton CE, Mwangi MM, Dewell S, Papavasiliou FN. Transcriptome-wide sequencing reveals numerous APOBEC1 mRNA-editing targets in transcript 3′ UTRs. Nat Struct Mol Biol 2011; 18(2): 230-236.

[8]   Azad M.T.A., Bhakta S., and Tsukahara T. Site-directed RNA editing by adenosine deaminase acting on RNA (ADAR1) for correction of the genetic code in gene therapy. Gene therapy, 24(12): 779–786, (2017).

[9]   Montiel-Gonzalez M.F., Vallecillo-Viejo I., Yudowski G.A., and Rosenthal J.J.C. Correction of mutations within the cystic fibrosis transmembrane conductance regulator by site-directed RNA editing. Proc. Natl. Acad. Sci. USA, 110(45), 18285–18290, (2013).

[10]   Stafforst T., and Schneider M.F. An RNA Deaminase conjugates electively repairs point mutations. Angew. Chem. Int. Ed, 51(44): 11166–11169, (2012).

[11]   Azad M.T.A, Qulsum U., and Tsukahara T. Comparative Activity of Adenosine Deaminase Acting on RNA (ADARs) Isoforms for Correction of Genetic Code in Gene Therapy. Current Gene Therapy, 19(1): 31 – 39, (2019).

[12]   Vogel P., Schneider M.F., Wettengel J., and Stafforst T. Improving site-directed RNA editing in vitro and in cell culture by chemical modification of the guideRNA. Angew. Chem. Int. Ed, 53(24): 6267–6271, (2014).

[13]   Hanswillemenke A., Kuzdere T., Vogel P., Jékely G., and Stafforst T. Site-directed RNA editing in vivo can be triggered by the light-driven assembly of an artificial riboprotein. J. Am. Chem. Soc, 137(50): 15875–15881, (2015).

[14]   Montiel-Gonzalez MF, Vallecillo-Viejo I, Yudowski GA, Rosenthal JJ. Correction of mutations within the cystic fibrosis transmembrane conductance regulator by site-directed RNA editing. Proc Natl Acad Sci 2013; 110(45): 18285-18290.

[15]   Vu LT, Nguyen TTK, Thoufic AAM, Suzuki H, Tsukahara T. Chemical RNA editing for genetic restoration: the relationship between the structure and deamination efficiency of carboxyvinyldeoxyuridine oligodeoxynucleotides. Chem Biol Drug Des 2016; 87(4): 583-593.

[16]   Vu LT, Nguyen TTK, Alam S, Sakamoto T, Fujimoto K, Suzuki H et al. Changing blue fluorescent protein to green fluorescent protein using chemical RNA editing as a novel strategy in genetic restoration. Chem Biol Drug Des 2015; 86(5): 1242- 1252.

[17]   Wettengel J, Reautschnig P, Geisler S, Kahle PJ, Stafforst T. Harnessing human ADAR2 for RNA repair–recoding a PINK1 mutation rescues mitophagy. Nucleic Acids Res 2016; 45: 1-12.

[18]   Keryer-Bibens C, Barreau C, Osborne HB. Tethering of proteins to RNAs by bacteriophage proteins. Biol Cell 2008; 100(2): 125-138.

[19]   Buxbaum AR, Haimovich G, Singer RH. In the right place at the right time: visualizing and understanding mRNA localization. Nat Rev Mol Cell Biol 2015; 16(2): 95-109.

[20]   Rinkevich FD, Schweitzer PA, Scott JG. Antisense sequencing improves the accuracy and precision of A-to-I editing measurements using the peak height ratio method. BMC Res Notes. 2012 Jan 24;5:63.

[21]   Montiel-González MF, Vallecillo-Viejo IC, Rosenthal JJ. An efficient system for selectively altering genetic information within mRNAs. Nucleic Acids Res. 2016 Dec 1;44(21):e157.

[22]   Collier B, Gorgoni B, Loveridge C, Cooke HJ, Gray NK. The DAZL family proteins are PABP-binding proteins that regulate translation in germ cells. EMBO J. 2005 Jul 20;24(14):2656-66.

[23] Barreau C, Watrin T, Beverley Osborne H, Paillard L. Protein expression is increased by a class III AU-rich element and tethered CUG-BP1. Biochem Biophys Res Commun. 2006 Sep 1;347(3):723-30.

[24]   Heinrich S, Sidler CL, Azzalin CM, Weis K. Stem-loop RNA labeling can affect nuclear and cytoplasmic mRNA processing. RNA. 2017 Feb;23(2):134-141.

[25]   Fukuda M, Umeno H, Nose K, Nishitarumizu A, Noguchi R, Nakagawa H. Construction of a guide-RNA for site-directed RNA mutagenesis utilising intracellular A-to-I RNA editing. Sci Rep. 2017 Feb 2;7:41478.

[26]   M.T.A. Azad, S. Bhakta, T. Tsukahara, Site-directed RNA editing by adenosine deaminase acting on RNA for correction of the genetic code in gene therapy, Gene Ther. 24 (12) (2017) 779.

Chapter3 Improvement of MS2-ADAR system for site-directed RNA editing

3.1 Introduction

Single amino acid substitutions resulting from point mutations can cause tumor-specific antigens, contributing to the development of cancer [1]. Genetic diseases such as cystic fibrosis [2], chronic granulomatous disease [3], hepatic lipase deficiency [4], and hemophilia A are also caused by point mutations. To address these diseases, genetic engineering technologies have been developed to control the expression and activities of intracellular target genes [5]. DNA editing is one of the genetic engineering methods that can be used to create site-specific double-strand DNA breaks (DSBs) followed by endogenous repair using either error-prone non-homologous end-joining (NHEJ) or error-free homology-directed repair (HDR) pathways [6-8]. However, these DNA editing technologies may also lead to the insertion and deletion mutations bridging the break site, which can cause cell death or oncogenic transformation [9-11].

One type of DNA editing, called base editing, allows for precise and irreversible conversion of one base pair to another at a specific genomic location without the need for double-stranded DNA breaks. The most commonly used third-generation base editors (BE3) use a deaminase-Cas9 fusion protein to change cytidine to uridine in target DNA without inducing a double-strand break [12-15]. However, as this editing affects genomic DNA, there is a risk of it being passed on to offspring during meiosis, potentially leading to unintended consequences. Off-target mutations can also occur, which can affect the

function or regulation of non-targeted genes. Additionally, larger structural alterations to the genome's sequence at the targeted editing location are also a concern [35].

RNA editing may be considered a safer alternative to DNA editing. This is because RNA editing involves post-transcriptional modifications, which allow for the production of multiple protein variants from a single gene [16]. Unlike DNA, RNA undergoes degradation after translation, meaning that any mutations resulting from RNA editing will not be passed down to future generations [17-19]. Furthermore, since mature mRNA lacks introns, intron editing is not a concern.

Different methodologies have been developed for site-directed RNA editing (SDRE) to achieve gene editing at the RNA level, particularly for adenosine (A)-to-inosine (I) RNA editing. Some of the methods include SNAP-tag [20], λN protein [21], MS2 system [22], and Cas13 [23]. Although these technologies achieve SDRE in different ways, they all follow the same principles.

In mammalian cells, adenosine (A) to inosine (I) and cytosine (C) to uridine (U) conversions occur in both coding and non-coding sequences [24-26]. In human cells, the apolipoprotein B mRNA editing enzyme (APOBEC1) was first identified to convert cytidine (C) to uridine (U) in apolipoprotein B (APOB) transcripts [27]. Meanwhile, the adenosine deaminase acting on RNA (ADAR) family mediates the conversion of A-to-I, with inosine serving as guanosine during translation [28, 29]. Recently, it has been reported that ADAR2 can be utilized for cytidine (C) to uridine (U) editing with mutation types [30].

In previous studies, we optimized the MS2 system for site-directed RNA editing, building on the λN-BoxB system [31]. To achieve this, we replaced the original CMV promoter with the U6 promoter for guide RNA expression, and incorporated a flexible

linker for the fusion protein. We also reduced the number of MS2 stem-loop RNA from 6 copies to 2 copies and positioned the guide RNA between the two stem-loops. Our optimization resulted in significantly improved editing efficiency compared to the previous version of the MS2 system [22].

However, previous studies have not compared the editing efficiency of the MS2 system when the number of stem-loops or the mismatch base is changed. Therefore, in this study, we investigated the impact of these factors on editing efficiency. We also compared the editing efficiency of different guide RNAs and their expression levels. Improving the editing efficiency of the system is crucial for its practical use.

3.2 MATERIALS AND METHOD

3.2.1 Plasmid construction

We used EGFP W58X as the target gene for SDRE, which was prepared in the previous study [22]. The expression vector for MS2 coat protein (MCP) and ADAR1 deaminase domain (ADAR1 DD) was also prepared in the previous study [31].

In this study, MCP harbored the N55K mutation at amino acid position 55 because it has a higher binding affinity to MS2 RNA than wild-type MCP [32]. The origin guide construct had prepared in the previous study [31]. The human U6 promoter (hU6 promoter) was substituted for the CMV IE94 promoter in pCS2+ (Addgene), and the guide RNA was put downstream of the hU6 promoter. The guide RNA sequence (5'-gaacatgaggatcacccatgtc**tgggccagggcacgggcagct**aacatgaggatcacccatgtctttt-3'), the sequence underlined represents the MS2 stem-loop RNA, the bold line section represents

antisense RNA. In the following, guide RNAs are distinguished by the numbers of MS2 stem-loop RNAs on the 5'side of antisense RNA and the number of MS2 stem-loop RNAs on the 3'side of antisense RNA, and the origin guide RNA construct is referred to as (1-1). For this research, first, we prepared three types of constructs with different numbers of MS2 stem-loop RNAs based on the (1-1) guide RNA constructs. Downstream of U6 promoter MS2 stem-loop RNA – antisense RNA – (MS2 stem-loop RNA)$_2$ (1-2), (MS2 stem-loop RNA)$_2$ - antisense RNA - MS2 stem-loop RNA (2-1), (MS2 stem-loop RNA)$_2$ - antisense RNA – (MS2 stem-loop RNA)$_2$ (2-2). Second, in order to examine each type of base pair in mismatches will affect the editing efficiency or not, use the (1-1) guide RNA construct as backbone, replaced the cytosine (C) that in the mismatch position into adenosine (A), uridine (U) or guanosine (G). These six types of guide RNA construct were used for this research. Primers for construction was shown in supplementary Table1.

Plasmid used in this article were transformed into Escherichia coli DH5α competent cells (TaKaRa Bio, Japan), then positive clones were cultured by 5mL LB medium with ampicillin at 37°C for 18 hours shaking and extracted by using NucleoBond Xtra Midi (MACHEREY-NAGEL, Germany). The concentration of isolated plasmid was determined using a NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific, MA, USA).

3.2.2 Cell Culture

HEK293T cells (from RIKEN BRO CELL BANK) were maintained in Dulbecco's modified Eagle's medium (Nacalai Tesque, Kyoto, Japan) with 10% fetal bovine serum (Thermo Fisher Scientific) under 5% CO_2 at 37°C. After at least three passages from frozen stocks, cells were employed in investigations.

3.2.3 Transfection

$2.0 \sim 3.5 \times 10^5$ cells per well were seeded in 24-well culture plates (TrueLine) with 500µL DMEM with 10% FBS, grown for 24 hours to 70% confluence, and then subjected to transfection. Before transfection, 250µL medium was removed so that the volume per well was 250µL, and 250µL fresh DMEM with 10% FBS was added. Next, 50µL Opti-MEM (Gibco), 2.5µL PEI MAX (Polysciences, 1µg/µL), and 10ng of EGFP/EGFP W58X plasmid DNA, 100ng of MS2-ADAR1 plasmid DNA, the amount of guide RNA plasmid DNA was shown in Supplementary Table 4 and 5. were mixed, incubated for 20 minutes at room temperature, then added to each well. 24 hours after transfection, 300µL of medium was removed from each well, and 300µL fresh DMEM with 10%FBS was added.

3.2.4 Cell observation

48 hours after transfection, cells were observed on a Keyence Biozero-800 (Itasca, IL, USA) and BZ-X800 (Keyence Co., Ltd., Osaka, Japan) fluorescence microscope under standard conditions during observation. The fluorescence intensity was measured by ImageJ software (NIH, MD, USA).

3.2.5 RNA extraction and complementary DNA synthesis

Total RNA was extracted from transfected cells by the TRIzol reagent (Thermo Fisher Scientific) at 200µL per well according to the manufacturer's protocol. Extracted RNA concentration was measured by Nano Drop. Then, total RNA was treated by RNase-free Recombinant DNase I (TaKaRa) according to the manufacturer's protocol. Using 300-500ng of total RNA treated with Recombinant DNase I, cDNA was synthesized with

50 units of ReverTra Ace (TOYOBO, Osaka) by incubating for 30 minutes at 42℃, then incubating for 5 minutes at 90℃.

3.2.6 Determination of editing efficiency

To confirm the restoration of the target sequence, PCR was performed with Go Taq Flexi DNA polymerase (Promega) and 50ng of cDNA as a template to obtain the DNA region containing the target base in EGFP W58X. PCR products were purified by using NucleoSpin Gel and PCR Clean-up (MACHEREY-ANGEL).

Purified products Sanger sequencing was performed by Eurofins Genomics (Tokyo, Japan). The sequencing primer was shown in supplementary Table3. Raw sequence data were analyzed with the SnapGene Viewer (USA). When the edited and unedited products were present together, a dual peak G (edited) and A (unedited) was observed at the target site. Peak height was measured using the ImageJ software (NIH, MD, USA). Editing was quantified based on maximum peak height ratio of the edited and unedited products $\left\{100\% \times \left(\frac{G\ height}{A\ height\ +\ G\ height}\right)\right\}$ [36].

3.2.7 Determination of guide RNA expression level

For confirming each type guide RNA expression level, we perform the qPCR with specific primers. Primer used in this study are shown in Supplementary Table2. Each type guide cDNA is obtained by reverse transcription using random hexamer primers (Thermo Scientific, Cat#SO142). Real-time PCR was performed using a 20μL reaction by TB Green *Premix EX Taq* II (TaKaRa, Cat#RR820A). The instrumentation used was the Mx-3000P (Agilent Technologies). Initial denaturation was at 95℃ for 30 seconds followed by 50cycles of 95℃ denaturation for 5 seconds, and 50℃ annealing for 15 seconds, 72℃

extension for 15 seconds. Data were collected during the annealing and extension step. We used cDNA of human ACTB as interference.

3.3 Result

3.3.1 Examination of the effect of the number of MS2 stem-loop RNA on editing efficiency

In the previous research, the (1-1) guide RNA had shown high editing efficiency [31]. In this part, we examined the effect of the number of MS2 stem-loop RNA on RNA editing efficiency.

We transfected 390ng of each type of guide RNA construct with other 2 factors' construct into HEK 293T cells, 48 hours later we observed the fluorescence signal (**Fig. 1A**). From the fluorescence micrographs, the fluorescence intensity was measured as 6.06 Arbitrary Units (AU) in (1-1), followed by 4.27 AU in (2-2). 3.78 AU in (1-2) and 3.09 AU in (2-1) by ImageJ.

Then, we confirmed the editing efficiency by Sanger sequence (**Fig. 1B and C**), and each type of guide RNA expression level by qPCR (**Fig. 1D**). The editing efficiencies of 1-1 stem-loop, 1-2 stem-loop, 2-1 stem-loop and 2-2 stem-loop after 48 hours were 51.8%, 30.7%, 33.3% and 39.1%, respectively (**Fig. 1C**).

From the qPCR result, we found that the expression levels of different guide RNAs were different. The expression level of the guide RNA 1-1 is considered 1. The expression level of 1-2 stem-loop, 2-1 stem-loop and 2-2 stem-loop were 1.6, 1.3 and 2.1 (**Fig. 1D**).

A   Fluorescence images of transfected cells
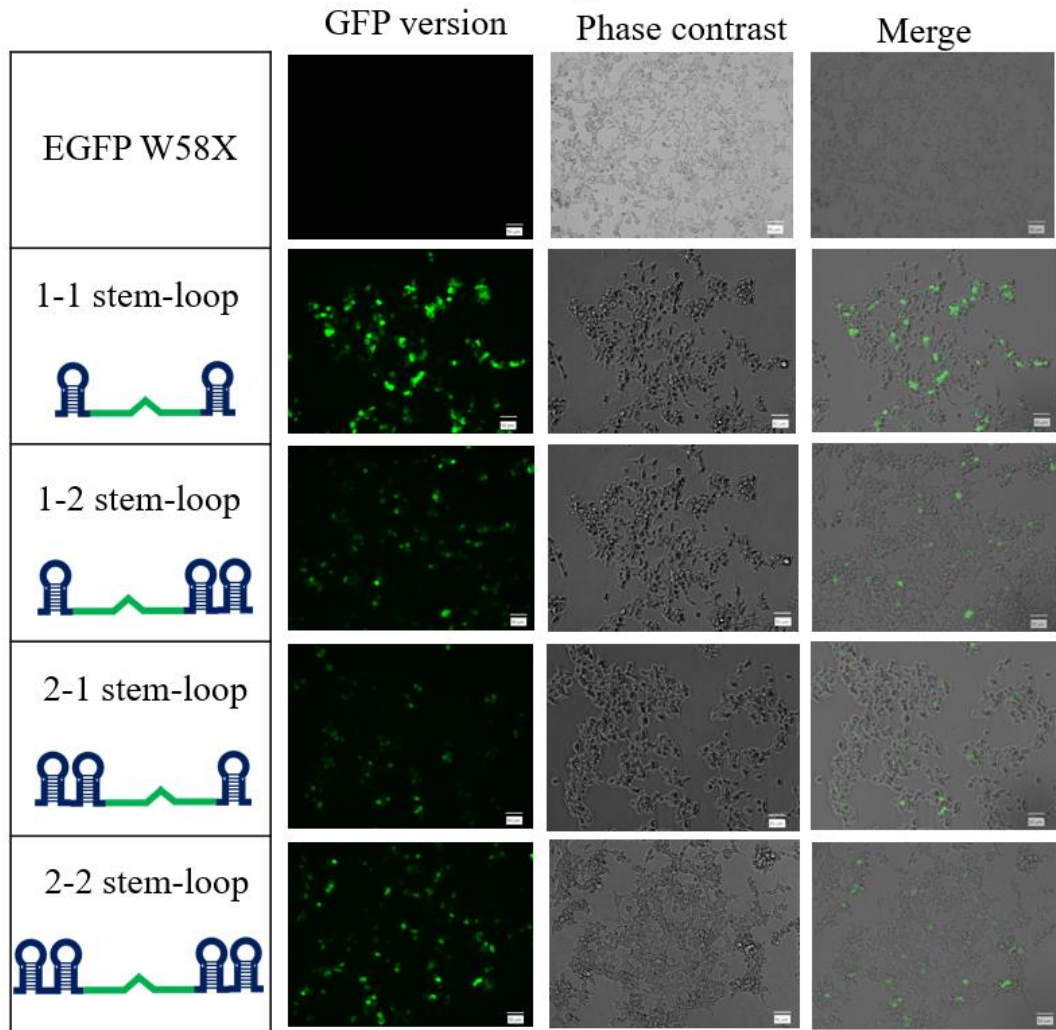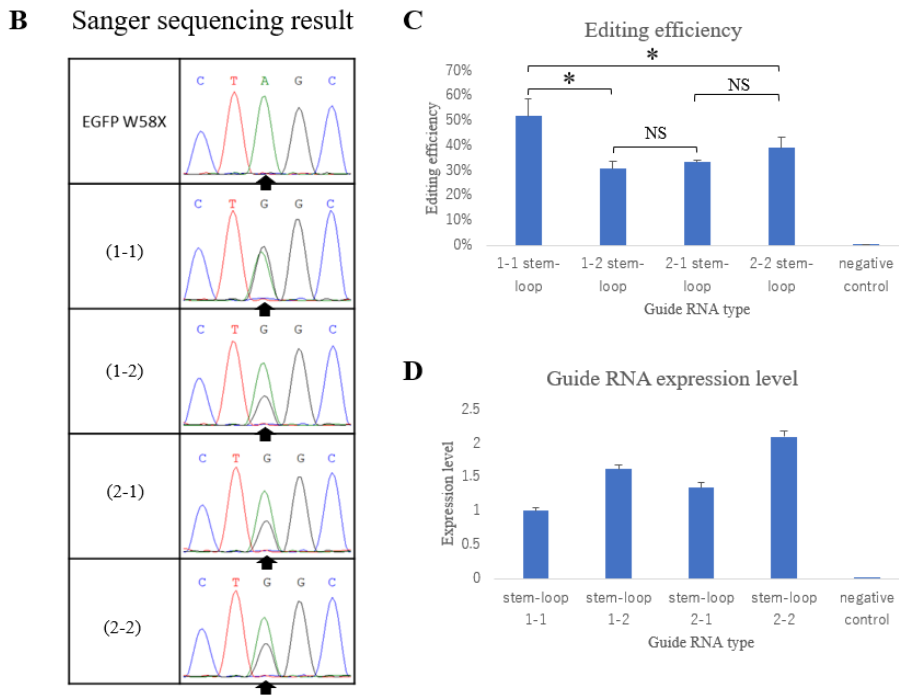
Fig. (1). A: Fluorescence micrographs of HEK293T were obtained 48 hours after transfection. The images on the left panel are fluorescence images, the middle panel is phase contrast, and the right panel is merged. Left panel of the fluorescence images is the schematic of guide RNA. **B:** Result of Sanger sequencing. Black arrows indicate the target base for EGFP W58X. **C:** Bar graph showing the editing efficiency (%) calculated from the sequencing results using the peak height ratio method. Mean and standard error of the mean (SEM) (n=3). EGFP W58X (negative control) editing level is considered 0%. Asterisks indicate a statistically significant difference ($p < 0.05$). NS=not significant. **D:** Bar graph showing each type of guide RNA expression level. The guide RNA 1-1 expression level is considered 1. Mean and standard error of the mean (SEM) (n=3).

3.3.2 Editing efficiency evaluation of the different number of MS2 stem-loop guide RNA
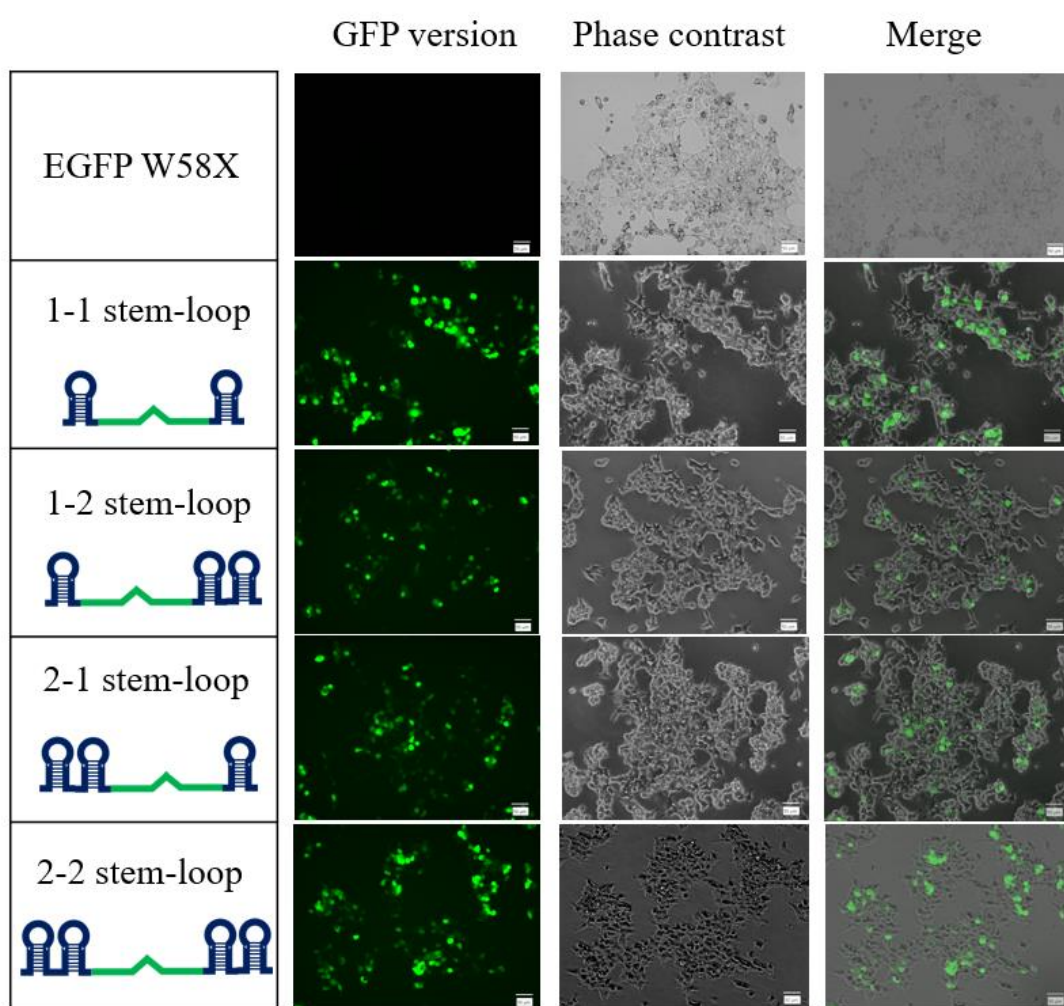
Since the molar ratio of each guide RNA to the target gene is different, it will also affect the evaluation of editing efficiency and cannot be compared, so we try to optimize the amount of plasmids for transfection.

We calculated the average molar ratio of guide RNA to target gene except for the negative control and optimize the amount of each type of guide RNA construct for transfection. Then, we transfected the 3 factors constructs into HEK293T cells and observed the fluorescence signal 48 hours later (**Fig. 2A**). From the fluorescence micrographs, the fluorescence intensity was measured as 10.30 AU in (1-1) and 10.28 AU in (2-2), 8.19 AU in (1-2) and 8.62 AU in (2-1) by ImageJ.

The editing efficiency and guide RNA expression level also had been calculated and shown as a bar graph (**Fig. 2C and D**). After optimization, the editing efficiencies of 1-1 stem-loop, 1-2 stem-loop, 2-1 stem-loop and 2-2 stem-loop after 48 hours were 53.3%, 33.3%, 33.8% and 54.8%, respectively (**Fig. 2C**).

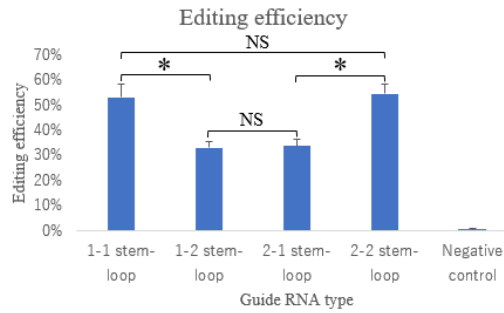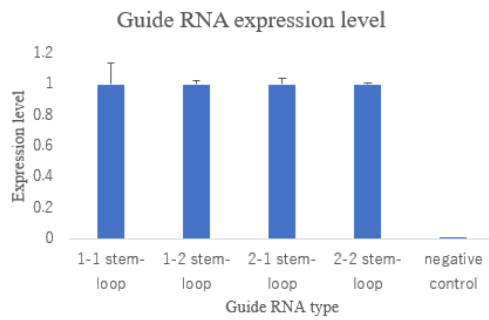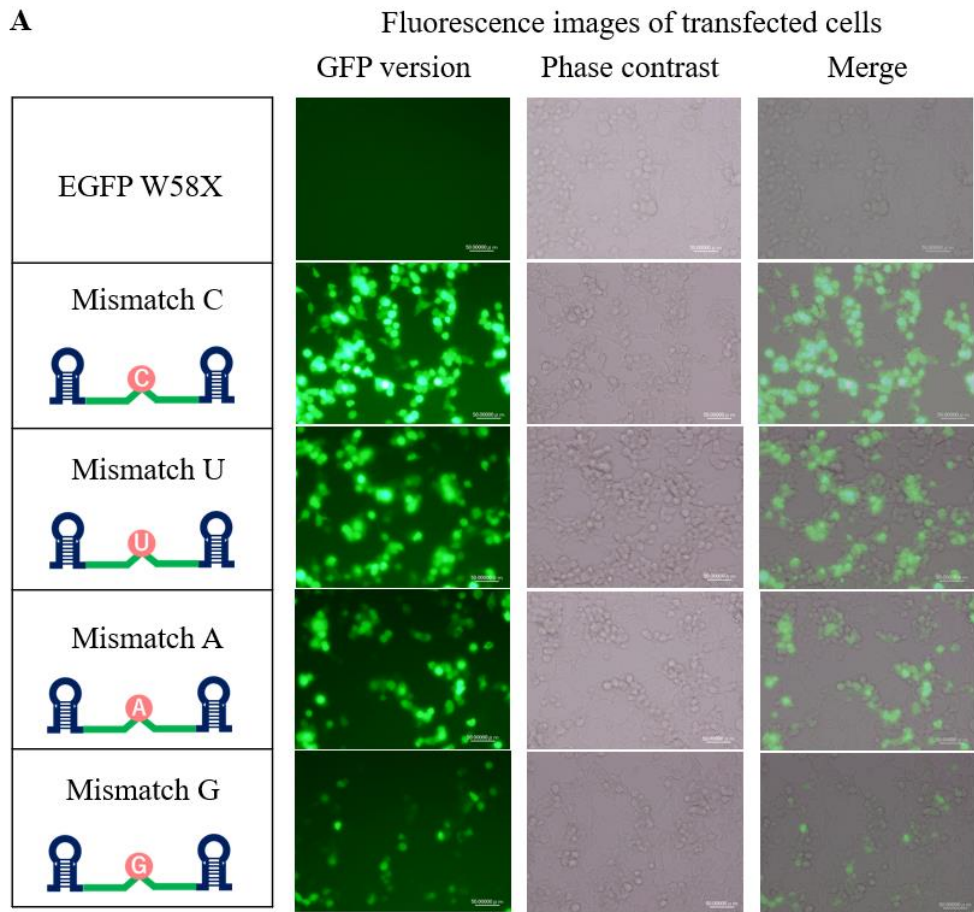**A**   Fluorescence images of transfected cells

Fig. (2). A: Fluorescence micrographs of HEK293T were obtained at 48 hours after transfection. The images on the left panel are fluorescence images, middle panel are phase contrast, and right panel are merged. Left panel of the fluorescence images is the schematic of guide RNA. B: Result of Sanger sequencing. Black arrows indicate the target base for EGFP W58X. C: Bar Graph showing the editing efficiency (%) calculated from the sequencing results using the peak height ratio method. Mean and standard error of the mean (SEM) (n=3). EGFP W58X (negative control) editing level is considered 0%. Asterisks indicate a statistically significant difference ($p < 0.05$). NS=not significant. D: Bar graph showing each type of guide RNA expression level. The guide RNA 1-1 expression level is considered 1. Mean and standard error of the mean (SEM) (n=3).

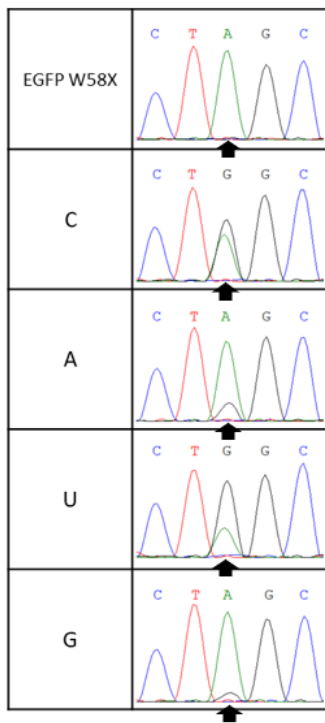### 3.3.3 Examination of the types of bases that pair with the target bases of SDRE

We transfected 390ng of each mismatch guide RNA construct with other 2 factors' construct into HEK 293T cells, 48 hours later we observed the fluorescence signal (**Fig. 3A**). The fluorescence intensity was measured as 51.60 AU in mismatch C, 33.17 AU in mismatch U, 23.40 AU in mismatch A, 18.38 AU in mismatch G.

Then, we performed the Sanger sequence (**Fig. 3B**) for calculating the editing efficiency (**Fig. 3C**) and confirmed each mismatch guide RNA expression level (**Fig. 3D**). Editing efficiency in mismatch C was 53.0%, mismatch A was 18.9%, mismatch U was 67.7%, and mismatch G was 9.1% (**Fig. 3C**).
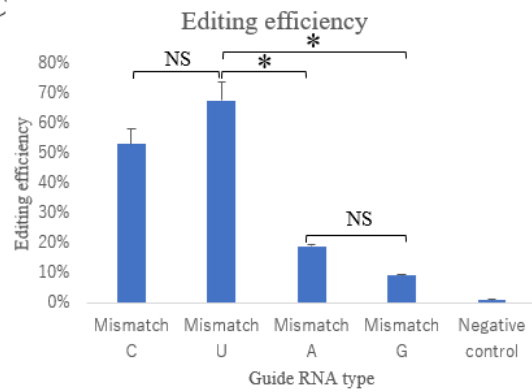
We also had confirmed that the expression level of each mismatch guide RNA by qPCR. We considered the expression level of mismatch C guide RNA is 1. The expression level of mismatch T, mismatch A and mismatch G were 0.39, 0.25 and 0.19 (**Fig. 3D**).

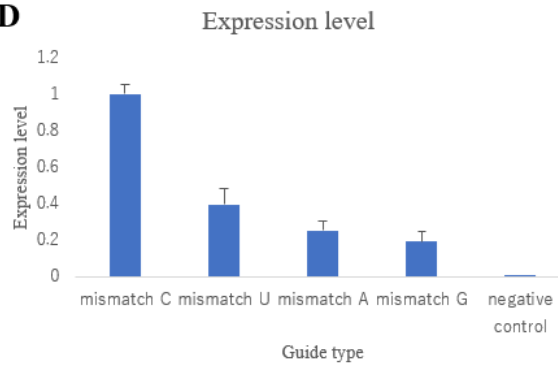**A** Fluorescence images of transfected cells

B Sanger sequencing result

C

D

Fig. 3. A: Fluorescence micrographs of HEK293T were obtained at 48 hours after transfection. The images on the left panel are fluorescence images, middle panel are phase contrast, and right panel are merged. Left panel of the fluorescence images is the schematic of guide RNA. **B:** Result of Sanger sequencing. Black arrows indicate the target base for EGFP W58X. **C:** Bar Graph showing the editing efficiency (%) calculated from the sequencing results using the peak height ratio method. Mean and standard error of the mean (SEM) (n=3). EGFP W58X (negative control) editing level is considered 0%. Asterisks indicate a statistically significant difference ($p < 0.05$). NS=not significant. **D:** Bar graph showing each type of guide RNA expression level. The guide RNA mismatch C expression level is considered 1. Mean and standard error of the mean (SEM) (n=3).
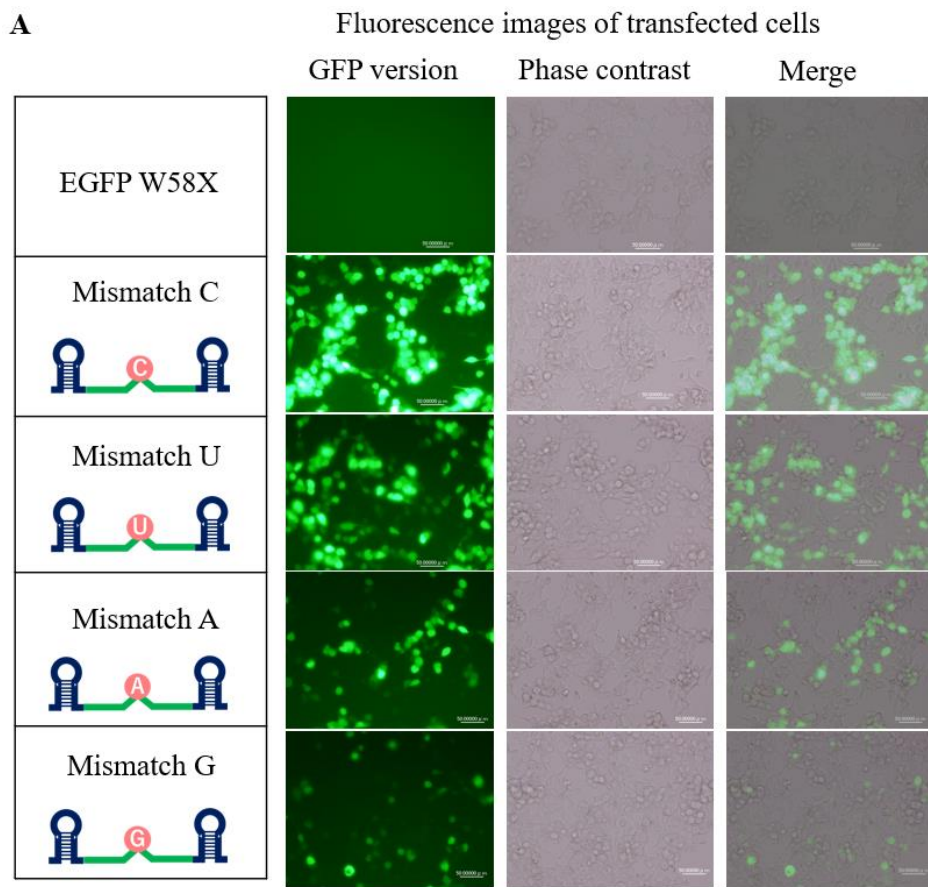
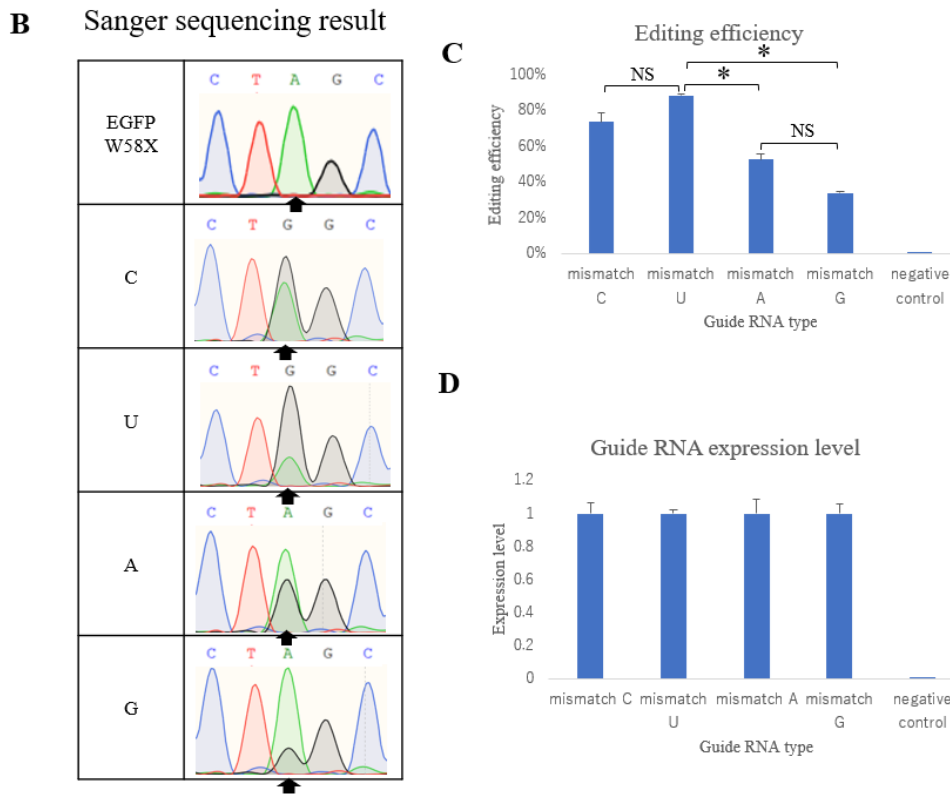# 3.3.4 Editing efficiency evaluation of the effect of each mismatch guide RNA on editing efficiency

We also adjusted the molar ratio of guide RNA to the target gene for transfection and compared the editing efficiency at the same molar ratio level.

Then, we also calculated the average of molar ratio of each mismatch guide RNA to the target gene and the amount of each mismatch guide RNA for transfection was optimized. We transfected the 3 factors constructs into HEK293T cells and the fluorescence signal was measured 48 hours later (**Fig. 4A**). The fluorescence intensity was measured as 59.31 AU in mismatch C, 33.07 AU in mismatch U, 21.12 AU in mismatch A and 15.84 AU in mismatch G.

The editing efficiency and guide RNA expression level also had been calculated and displayed as a bar graph (**Fig. 4C and D**). After optimization, editing efficiency in

mismatch C was 73.5%, mismatch A was 52.8%, mismatch U was 88.3%, and mismatch

G was 33.8% (**Fig. 4C**).

Fig. 4. A: Fluorescence micrographs of HEK293T were obtained at 48 hours after transfection. The images on the left panel are fluorescence images, middle panel are phase contrast, and right panel are merged. Left panel of the fluorescence images is the schematic of guide RNA. **B:** Result of Sanger sequencing. Black arrows indicate the target base for EGFP W58X. **C:** Bar Graph showing the editing efficiency (%) calculated from the sequencing results using the peak height ratio method. Mean and standard error of the mean (SEM) (n=3). EGFP W58X (negative control) editing level is considered 0%. Asterisks indicate a statistically significant difference ($p < 0.05$). NS=not significant. **D:** Bar graph showing each type of guide RNA expression level. The guide RNA mismatch C expression level is considered 1. Mean and standard error of the mean (SEM) (n=3).

3.4 Discussion

In this study, we aimed to investigate the effect of the number of stem-loops at the ends of the antisense RNA on editing efficiency (as illustrated in Fig. 1 and 2). To

maintain consistent guide RNA expression, we adjusted the number of constructs used for transfection and analyzed the results. Although changes in editing efficiency were difficult to discern through fluorescence observation, Sanger sequencing revealed significant differences. Notably, the 2-2 stem-loop configuration resulted in increased editing efficiency despite lower expression levels (as shown in Fig. 2C and D). However, certain configurations such as 1-2 stem-loop and 2-1 stem-loop did not demonstrate changes in editing efficiency even after optimizing the number of constructs used for transfection (as depicted in Fig. 1C and Fig. 2C). These results suggest that the MS2 SDRE system can achieve high editing efficiency when the number of stem-loops at both ends of the antisense RNA is equal [37].

In previous studies, it was reported that the molar ratio of guide RNA to target gene could affect editing efficiency, which is a crucial factor for gene therapy. For instance, one study using λN achieved a high editing efficiency of up to 57%, but the molar ratio of guide RNA expression plasmid to the target gene expression plasmid was 60:1, indicating that the guide RNA was 60-fold more abundant than the target gene [26]. Another study using Cas13 reported a molar ratio of 7.5:1 for the guide RNA expression plasmid to the target gene expression plasmid (RNA editing reporter) [23, 37].

In our experiment, we investigated the impact of the molar ratio of guide RNA expression plasmid to EGFP W58X target gene expression plasmid on editing efficiency. Before adjustment for transfection, the molar ratios for 1-1 stem-loop, 1-2 stem-loop, 2-1 stem-loop, and 2-2 stem-loop were 59:1, 36:1, 44:1, and 28:1, respectively. After adjustment, the molar ratio was reduced to 39:1. Surprisingly, even though the molar ratio was similar for 1-1 stem-loop and 2-2 stem-loop after adjustment, they still achieved similar editing efficiency (Fig. 2C). Additionally, the guide RNA expression level of 1-1

stem-loop and 2-2 stem-loop was also similar (Fig. 2D). This result is difficult to explain, but it may be related to the interaction between the MS2 coat protein and the stem-loop of the MS2 RNA. The unequal number of stem-loops at both ends of the antisense RNA may result in different distances between the MS2 coat protein bound to the MS2 RNA and the target site, thereby affecting the efficiency of RNA editing [37].

Next, we investigated whether different base mismatch pairings would affect editing efficiency. In previous studies, the highest editing efficiency was observed when the mismatch was C [33]. However, in contrast to those findings, our results showed that the highest editing efficiency was achieved with a mismatch of U. This is an intriguing result because the highest editing efficiency occurs when the bases are perfectly matched. The reason for this result is unclear, but it may be related to the structure of the ADAR1 catalytic domain. The perfectly matched double-stranded structure may be more conducive to the binding of the ADAR1 catalytic domain to the target gene, thereby improving the catalytic efficiency [37].

There are still many unanswered questions in this study, such as whether different mismatch bases will affect protein translation, and whether the number of stem-loops at both ends of the guide RNA affects editing efficiency when it is an odd or even number. Furthermore, it remains to be verified whether the results of this experiment can be applied to the MS2-APOBEC1 system [34, 37].

3.5 Conclusion

This study aimed to further verify the catalytic conditions of the MS2-ADAR1 system and optimize it for higher editing efficiency. The results showed that the editing efficiency could reach up to 80% by adjusting the conditions of guide RNA, which

suggests the potential of this system in future gene therapy applications. Further optimization of the MS2-ADAR1 system is possible based on these findings. Additionally, replacing ADAR1 with APOBEC1 can also achieve C to U catalysis, indicating that the MS2 system can have a wider range of RNA editing functions. However, there are still several unanswered questions, such as the effect of different mismatch bases on protein translation, the impact of odd or even number of stem-loops, and the transferability of the results to MS2-APOBEC1 system. Further research is needed to address these questions and expand our understanding of this system.

3.6 Reference

[1]   Matthias Leisegang; Boris Engels; Karin Schreiber; Poh Yin Yew; Kazuma Kiyotani; Christian Idel; Ainhoa Arina; Jaikumar Duraiswamy; Ralph R Weichselbaum; Wolfgang Uckert; Yusuke Nakamura; Hans Schreiber. Eradication of large solid tumors by gene therapy with a T-cell receptor targeting a single cancer-specific point mutation. *Clin Cancer Res* **2016**; *22*: 2734–2743.

[2]   C Férec 1; M P Audrezet; B Mercier; H Guillermit; P Moullier; I Quere; C Verlingue. Detection of over 98% cystic fibrosis mutations in a Celtic population. *Nat Genet*. **1992**; *1(3)*: 188–191.

[3]   Bolscher BG; M de Boer; A de Klein; Weening RS; Roos D. Point mutations in the beta-subunit of cytochrome b558 leading to X-linked chronic granulomatous disease. Blood. **1991**; *77*: 2482–2487

[4]   M Higuchi; H H Kazazian Jr; L Kasch, T C Warren; M J McGinniss, J A Phillips; C Kasper; R Janco; S E Antonarakis. Molecular characterization of severe hemophilia A suggests that about half the mutations are not within the

coding regions and splice junctions of the factor VIII gene. *Proc Natl Acad Sci USA*. **1991**; *88*: 7405–7409.

[5]   Matthias Heidenreich; Feng Zhang. Applications of CRISPR-Cas systems in neuroscience. *Nat Rev Neurosci*; **2016 Jan**; *17(1)*:36-44.

[6]   Maeder, M.L.; Gersbach, C.A. Genome-Editing Technologies for Gene and Cell Therapy. *Mol. Ther*. **2016**, *24*, 430–446.

[7]   Zhang, H.; Zhang, J.; Lang, Z.; Botella, J.R.; Zhu, J.-K. Genome Editing—Principles and Applications for Functional Genomics Research and Crop Improvement. *Crit. Rev. Plant Sci*. **2017**, *36*, 291–309.

[8]   Shan, Q.; Wang, Y.; Li, J.; Gao, C. Genome Editing in Rice and Wheat Using the CRISPR/Cas System. Nature protocols. **2014**, *9*, 2395–2410.

[9]   Zhang XH; Tee LY; Wang XG; Huang QS; Yang SH. Off-target effects in CRISPR/Cas9-mediated genome engineering. *Mol Ther Nucleic Acids*. **2015** Nov 17;*4(11)*: e264.

[10]   Cox, D.B.T.; Platt, R.J.; Zhang, F. Therapeutic Genome Editing: Prospects and Challenges. *Nat. Med*. **2015**, *21*, 121–131.

[11]   Zhang, F.; Wen, Y.; Guo, X. CRISPR/Cas9 for Genome Editing: Progress, Implications and Challenges. *Hum. Mol. Genet*. **2014**, *23*, R40–R46.

[12]   Lei S Qi; Matthew H Larson; Luke A Gilbert; Jennifer A Doudna; Jonathan S Weissman; Adam P Arkin; Wendell A Lim. Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. Cell **2013** Feb 28;*152(5)*:1173-83.

[13]   Martin Jinek; Krzysztof Chylinski; Ines Fonfara; Michael Hauer; Jennifer A Doudna; Emmanuelle Charpentier. A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science **2012** Aug 17;*337(6096)*:816-21.

[14]   Reuben S.Harris; Svend K.Petersen-Mahrt; Michael S.Neuberger. RNA Editing Enzyme APOBEC1 and Some of Its Homologs Can Act as DNA Mutators. Molecular Cell *10(5)*, **2002**, 1247-1253.

[15]   Gaudelli NM; Komor AC; Rees HA; Packer MS; Badran AH; Bryson DI; Liu DR. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. Nature. **2017**; *551*:464–471.

[16]   Maas S, Rich A. Changing genetic information through RNA editing. BioEssays **2000**; *22*: 790–802.

[17]   Paul Vogel; Thorsten Stafforst. Critical Review on Engineering Deaminases for Site-Directed RNA Editing. *Curr. Opin. Biotechnol*. **2019**, *55*, 74–80.

[18]   Montiel-Gonzalez; M.F. Quiroz; J.F.D. Rosenthal; J.J. Current Strategies for Site-Directed RNA Editing Using ADARs. Methods **2019**, *156*, 16–24.

[19]   Genghao Chen; Dhruva Katrekar; Prashant Mali. RNA-Guided Adenosine Deaminases: Advances and Challenges for Therapeutic RNA Editing. Biochemistry **2019**, *58*, 1947–1957.

[20]   Paul Vogel; Matin Moschref; Qin Li; Tobias Merkle; Karthika D. Selvasaravanan; Jin Billy Li; Thorsten Stafforst; Efficient and Precise Editing of

Endogenous Transcripts with SNAP-Tagged ADARs. Nature Methods **2018**, *15*, 535–538.

[21]   Isabel C Vallecillo-Viejo; Noa Liscovitch-Brauer; Maria Fernanda Montiel-Gonzalez; Eli Eisenberg; Joshua J C Rosenthal. Abundant Off-Target Edits from Site-Directed RNA Editing Can Be Reduced by Nuclear Localization of the Editing Enzyme. *RNA Biol*. **2018**, *15*, 104–114.

[22]   Azad MTA; Bhakta S; Tsukahara T; Site-directed RNA editing by adenosine deaminase acting on RNA for correction of the genetic code in gene therapy, Gene Therapy *24(12)*, (**2017**) 779.

[23]   Cox, D.B.; Gootenberg, J.S.; Abudayyeh, O.O.; Franklin, B.; Kellner, M.J.; Joung, J.; Zhang, F. RNA Editing with CRISPR-Cas13. Science **2017**, *358*, 1019–1027.

[24]   Nishikura K. A-to-I editing of coding and non-coding RNAs by ADARs. *Nat Rev Mol Cell Biol* **2016**, *17*: 83–96

[25]   Slotkin W; Nishikura K. Adenosine-to-inosine RNA editing and human disease. *Genome Med* **2013**, *5*: 1–13.

[26]   Rosenberg BR; Hamilton CE; Mwangi MM; Dewell S; Papavasiliou FN. Transcriptome-wide sequencing reveals numerous APOBEC1 mRNA-editing targets in transcript 3′ UTRs. *Nat Struct Mol Biol* **2011**, *18*: 230–23

[27]   Powell LM; Wallis SC; Pease RJ; Edwards YH; Knott TJ; Scott J. A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine. Cell, **1987**, *50(6)*:831–40.

[28]   Keegan LP; Gallo A; O'Connell M a. The many roles of an RNA editor. *Nat Rev Genet*, **2001**, *2(11)*: 869–78.

[29]   Maas S, Rich A, Nishikura K. A-to-I RNA editing: Recent news and Residual Mysteries. *J Biol Chem* **2003**, *278*: 1391–1394.

[30]   Abudayyeh, O.O.; Gootenberg, J.S.; Franklin, B.; Koob, J.; Kellner, M.J.; Ladha, A.; Joung, J.; Kirchgatterer, P.; Cox, D.B.; Zhang, F. A Cytosine Deaminase for Programmable Single-Base RNA Editing. Science **2019**, *365*, 382–386.

[31]   Tetsuto   Tohama,   Matomo   Sakari   and   Toshifumi   Tsukahara. Development of a Single Construct System for Site-Directed RNA Editing Using MS2-ADAR. International Journal of Molecular Sciences **2020**; *21(14)*: 4943.

[32]   F Lim; M Spingola; D S Peabody. Altering the RNA binding specificity of a translational repressor. *J Biol Chem*. **1994** ;*269(12)*:9006-10.

[33]   Liang Qu; Zongyi Yi; Shiyou Zhu; Chunhui Wang; Zhongzheng Cao; Zhuo Zhou; Pengfei Yuan; Ying Yu; Feng Tian; Zhiheng Liu; Ying Bao; Yanxia Zhao & Wensheng Wei. Programmable RNA editing by recruiting endogenous ADAR using engineered RNAs. Nature Biotechnology, **2019**, *37*, 1059–1069.

[34]   Sonali Bhakta; Matomo Sakari; Toshifumi Tsukahara. RNA editing of BFP, a point mutant of GFP, using artificial APOBEC1 deaminase to restore the genetic code. *Sci Rep*. **2020**; *10*: 17304.

[35]   Ida Höijer; Anastasia Emmanouilidou; Rebecka Östlund; Robin van Schendel; Selma Bozorgpana; Marcel Tijsterman; Lars Feuk; Ulf Gyllensten; Marcel den Hoed; Adam Ameur. CRISPR-Cas9 induces large structural variants at

on-target and off-target sites in vivo that segregate across generations. *Nature Communications;* **2022**; *13*; 627.

[36]   Julie M. Eggington; Tom Greene; Brenda L. Bass. Predicting sites of ADAR editing in double-stranded RNA. *Nat. Commum*; **2011**, *2*, 1-9

[37]   Li, J.; Oonishi, T.; Fan, G.; Sakari, M.; Tsukahara, T. Increasing the Editing Efficiency of the MS2-ADAR System for Site-Directed RNA Editing. Appl. Sci. 2023, 13, 2383. https://doi.org/10.3390/app13042383

Chapter4: Programmable C-to-U RNA editing using human APOBEC3A and APOBEC3G deaminase

4.1 Introduction

RNA editing is a post-transcriptional process that modifies transcript sequences without altering the encoding DNA sequence [1]. A-to-I and C-to-U conversions are the most widely used strategies for site-directed RNA editing (SDRE) in recent years [2-6]. Previous studies have used different methods for performing adenine deamination, including SNAP-tag [7], λN protein [8], Cas13 [9], and the MS2 system [10]. However, all of these methods allow for only single-base precision in A-to-I (G) conversion [11]. The development of technologies for precise RNA editing of cytidine-to-uridine conversions would greatly expand the range of addressable disease mutations and protein modifications [12].

In recent years, Zhang's lab has developed the RESCUE system for performing C-to-U conversions [12]. This system involves coupling the RNA-targeting CRISPR-Cas13b to the mutant human ADAR2 deaminase domain (dCas13b). Although the RESCUE system is capable of introducing the A-to-I off-target conversion in addition to the C-to-U conversion at the target site, its efficiency for C-to-U editing is still limited.

To address this issue, the RESCUE-S system was developed to reduce A-to-I off-targets, but this also led to a decrease in C-to-U editing efficiency [13].

In mammalian cells, C-to-U RNA editing was discovered 35 years ago as the molecular basis for human intestinal apolipoprotein B 48 (apoB48) production [14-16]. Members of the apolipoprotein B-editing catalytic polypeptide family, including mammalian CDA proteins and activation-induced deaminase (AID), possess the CDA motif necessary for hydrolytic deamination of C-to-U [17]. The APOBEC3 (A3) family of cytidine deaminases in the APOBEC family plays a crucial role in the vertebrate innate immune system by limiting endogenous retroelements and foreign viruses [18-20]. Previous studies by the Baysal lab demonstrated that APOBEC3A and APOBEC3G can introduce C-to-U deamination at a specific site in single-stranded (ss) RNA [21, 22]. In 2020, the CURE system was developed for C-to-U RNA editing by fusing the human APOBEC3A to dCas13b [13].

In 2020, we utilized the MS2 system to achieve site-directed C-to-U deamination with an artificial APOBEC1, but the editing efficiency was only up to 21%, which limits its practical applications. In this article, we aim to improve the editing efficiencies of the MS2 system by using human APOBEC3A (A3A) and APOBEC3G (A3G). A3A has a base preference of 5'-TC, while A3G prefers 5'-CC. To enhance A3G's preference for 5'-TC, we introduced the D317W mutation into A3G [24]. We also used a well-established single amino acid substitution in enhanced Green Fluorescent Protein (eGFP) that leads to an open reading frame associated with a transition to Blue Fluorescent Protein (BFP), as described in previous studies [23, 25, 26].

In this article, we also investigate the effect of different loop lengths of the target on editing efficiency, in order to ensure the practical usability of the system [27].

4.2 Materials and Methods

4.2.1 Plasmid Construction

Reporter gene plasmid that expresses BFP (Addgene, MA, USA) was used as a reporter gene in this study [23]. The backbone vector was pcDNA 3. For the base preference of A3A and A3G, we introduced 5'- ACT<u>C</u>AC mutagenesis into the expression vector, underlined letter C indicates the target site [26].

DNA sequences of all primers used in this paper are listed in the Supplementary. For expression of MS2 coat protein (MCP) -GS linker-XTEN linker-human APOBEC3A/APOBEC3G, we designed three pairs of primers for cloning and construction (Supplementary). PCR was performed using KOD-Plus-Neo (TOYOBO) according to the manufacturer's instructions. Fusion protein plasmids were constructed using NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs). APOBEC3A deaminase gene was cloned from human lung cDNA. APOBEC3G deaminase gene was cloned from human cDNA clone vector (HGY096998, RIKEN BRC DNA BANK, Ibaraki, Japan). Wild-type A3A and A3G cDNA sequences used in this study match the NCBI reference sequences NM_145699.2 and NM_021822.1. Deaminase genes were cloned into pCS2-MCP-GS linker-XTEN linker backbone that had been used in the previous study [28]. For the D317W mutation, we used KOD-Plus-Neo (TOYOBO) to perform the site-directed mutagenesis PCR with specific primers (Supplementary) and pCS2-MCP-GS linker-XTEN linker- APOBEC3G (natural type).

For guide RNA (gRNA) plasmid construction, the oligos listed in the Supplementary and were synthesized by Eurofins Genomics (Tokyo, Japan). The backbone plasmid PC0003 was constructed in the previous study [27]. The oligos annealing followed the

protocol for annealing oligonucleotides. PC0003 was digested by Bbs I (New England Biolabs) and ligated with the annealed oligonucleotides by Ligation high ver.2 (TOYOBO). All the plasmids were extracted by the NucleoSpin Plasmid Transfection-grade (MACHEREY-ANGEL, Düren, Germany).

4.2.2 Cell culture and Transfection

HEK 293T cells (From RIKEN BRC CELL BANK) were cultured at 37°C with 5% CO_2 in DMEM (Nacalai Tesque, Kyoto, Japan) containing high glucose, sodium pyruvate, and 10% fetal bovine serum Cells were passaged three times per week.

Transfections were performed with Polyethylemine Hydrochloride (PEI, Polysciences, Illinois, USA) in 12-well plates per manufacturer's instruction. Cells were plated into 12-well plates at 32×10^4 / well and transfected 16 h later. 1 ng Plasmid DNA was mixed with 5 µL PEI into 100 µL Opti-MEM (Invitrogen) and incubated for 20 min at the room temperature. The DNA : PEI mixture was then added dropwise into the wells. The medium would be removed 24 h after transfection. Cells were observed using BZ-8000 FLUORESCENCE MICROSCOPE (Keyence, Osaka, Japan) and FV1000D confocal laser-scanning microscope (Olympus, Tokyo, Japan) 48 h post-transfection.

4.2.3 Reverse transcription and PCR reaction

48 h after transfection, total RNA was extracted using TRIzol Reagent (Invitrogen). cDNA was synthesized from the extracted RNA using the ReverTra Ace (TOYOBO, Osaka, Japan).

To confirm the restoration of the editing, PCR was performed with GoTaq Flexi DNA Polymerase (Promega, Wisconsin, USA). The primer had been shown in the supplementary. The PCR products were purified using NucleoSpin Gel and PCR Clean-up (MACHEREY-ANGEL).

4.2.4 Determination of editing efficiency

After PCR product purification, the Sanger sequencing was performed by Eurofins Genomics (Tokyo, Japan). The sequencing primer was shown in the Supplementary. The raw sequencing data were analyzed using the Snap gene (Dotmatics, USA). When the edited and unedited products were presented together, a dual peak (C[unedited] and T [edited]) was observed at the target site. Following the precious works on calculation of editing efficiency from peak area and peak height, we also calculated by measuring the area and peak height using the ImageJ software (NIH, https://imagej.net/Citing).

$$\text{Considering Peak height:} \quad \frac{Peak\ height\ of\ T}{Peak\ height\ of\ T + Peak\ height\ of\ C} \times 100\%$$

4.2.5 Statistical Analysis

In all experiments for determining editing efficiency, three independent experiments were performed. Multiple comparisons were made by performing a paired t-test followed by the Bonferroni correction method.

4.3 Result

4.3.1 Engineering MS2-APOBEC3A system for specific C-to-U conversion

In the previous study, CURE system showed the possibility of RNA editing by human APOBEC3A link with CRISPR-Cas13b. We assumed it would be possible to fuse APOBEC3A to the MS2 coat protein. In this study, we didn't fuse nuclear export sequence (NES) or nuclear localization signal (NLS). Because previous research found that NLS would decrease editing efficiency [28]. We focused on the MS2-APOBEC3A for inducing C-to-U RNA editing and improving the editing efficiency.

We redesigned the reporter system based on the previous one (**Fig 1A**) [23]. Due to the base preference of APOBEC3A, we performed site-directed mutagenesis on the reporter vector (**Fig 1B**). The target cytosine had been marked in red.

Referring to the guide RNA (gRNA) structure of the CURE system, we also introduced a cassette for gRNA expression under the human U6 promoter into the 1-1 MS2 stem-loop expression vector to generate a similar guide RNA for MS2-APOBEC3A system (**Fig 1C**) [28]. To investigate whether stem-loop length affects editing efficiency, we designed six types of guide RNAs to induce loops of six lengths (**Fig 1D**).

A

Codon	Amino acid
CAC	Histidine

Codon	Amino acid
TAC	Tyrosine
TAT	Tyrosine
CAT	Histidine

B

CCACCCTGAC<u>CC</u>ACGGCGTG

Site-firected mutagenesis

CCACCCTGAC<u>T</u>CACGGCGTG

C

D

The loop length	Target sequence
20nt loop	CCACCCUGACUCACGGCGUG
18nt loop	CACCCUGACUCACGGCGU
16nt loop	ACCCUGACUCACGGCG
14nt loop	CCCUGACUCACGGC
12nt loop	CCUGACUCACGG
10nt loop	CUGACUCACG

Fig. 1: A florescent reporter detects RNA editing activity. **A.** GFP is expressed as a result of targeting codon 66 (CAC), which codes for histidine, and converting it to

tyrosine (codons UAU or UAC). Since His is represented by the codon conversion to

CAU, the protein continues to be BFP. **B.** Diagram of site-directed mutagenesis site.

Target cytosine (C) had been marked in red. **C.** BFP reporter base pair with guide RNA,

predicted to induce a 14 nucleotide (nt) loop encompassing ACU<u>C</u>, was tested together

with various controls. **D.** Six types of loop lengths and sequence for APOBEC3A RNA

editing.

After we co-transfected three factors into HEK 293T cells, we observed the

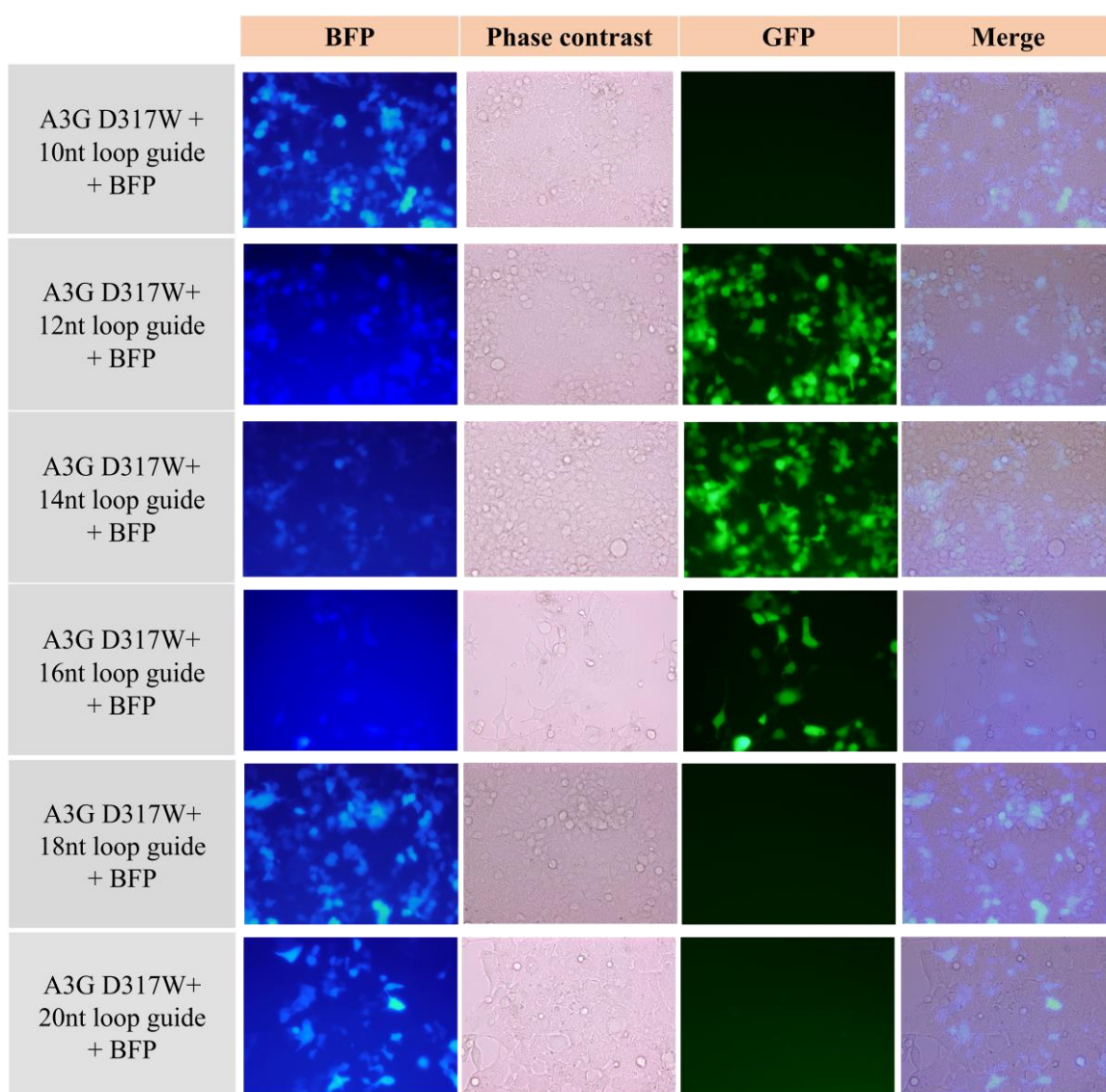fluorescent signal by Keyence BZ-8000 fluorescence microscopy (**Fig 2**).
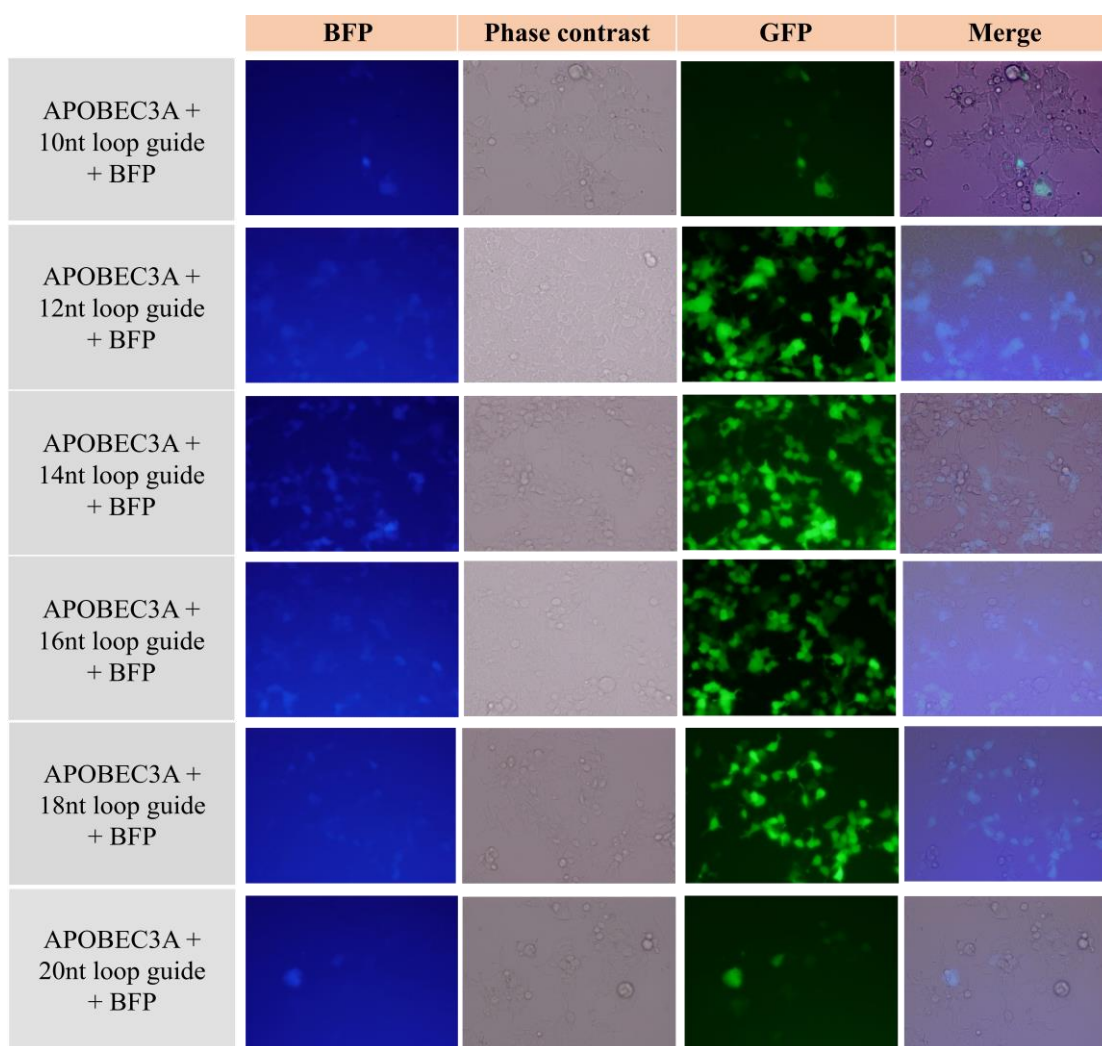
Fig 2. Fluorescence micrographs of HEK293T were obtained at 48h after transfection. The images on the left panel are BFP vision, the images on the second left panel are phase contrast, the images on the third left panel are GFP vision, and the right panel is merged.

From the figure 2, we found that the fluorescent signal of 10nt loop and 20nt were very weak. The fluorescent signal of 12nt loop and 14nt loop were stronger than 16nt loop and 18nt loop. We extracted the RNA from the transfected cells for confirming the editing efficiency of each condition (**Fig 3**). 14nt loop guide showed the highest editing efficiency

in all conditions. In the sanger sequencing result, we didn't detect any deamination signal of other C except the target C on the nucleotide loop (Fig 4).
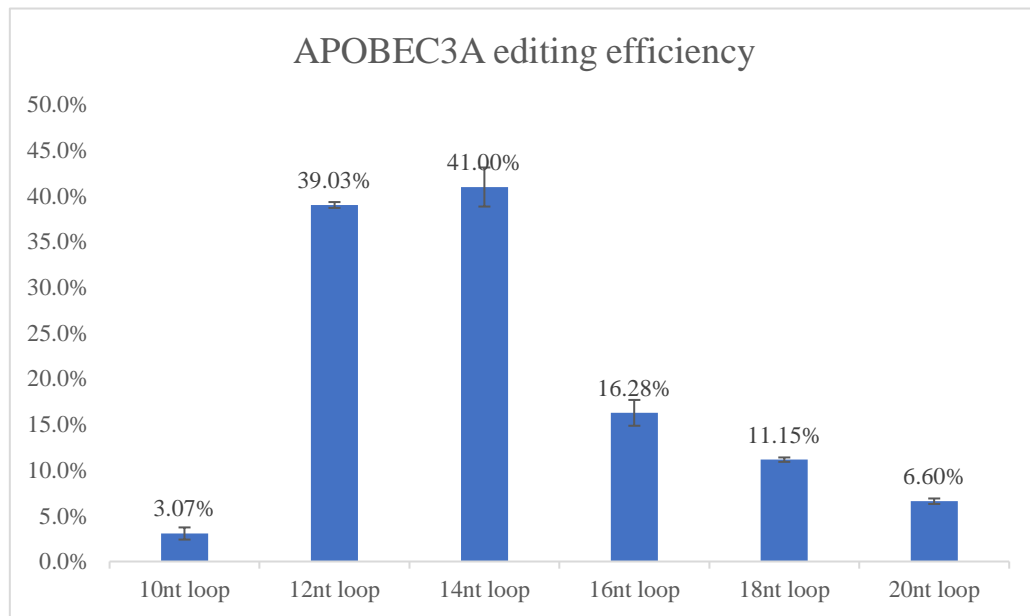
Fig 3: Bar graph showing the editing efficiency (%) calculated from the sequencing results using the peak height ratio method. Mean and standard error of the mean (SEM) (n=3).
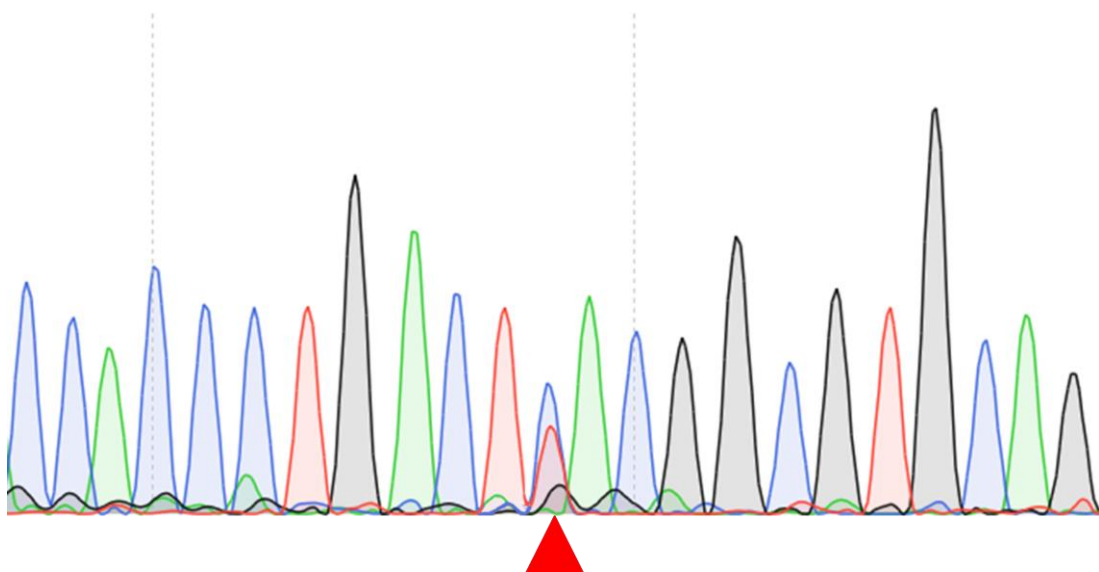
Fig4: Sanger sequencing result of MS2-APOBEC3A transfected with 14nt loop

guide RNA

4.3.2 Engineering MS2-APOBEC3G system for specific

C-to-U conversion

After the success of APOBEC3A, we tried to perform the same method on the APOBEC3G. Because APOBEC3G also can induce C-to-U RNA editing on the single strand RNA. We constructed the MS2-APOBEC3G by the same method, then co-transfected the guide RNA plasmids and BFP plasmid into HEK 293T. The observation was performed by Keyence BZ-8000 fluorescence microscopy.

	BFP	Phase contrast	GFP	Merge
APOBEC3G + 10nt loop guide + BFP				
APOBEC3G + 12nt loop guide + BFP				
APOBEC3G + 14nt loop guide + BFP				
APOBEC3G + 16nt loop guide + BFP				
APOBEC3G + 18nt loop guide + BFP				
APOBEC3G + 20nt loop guide + BFP				
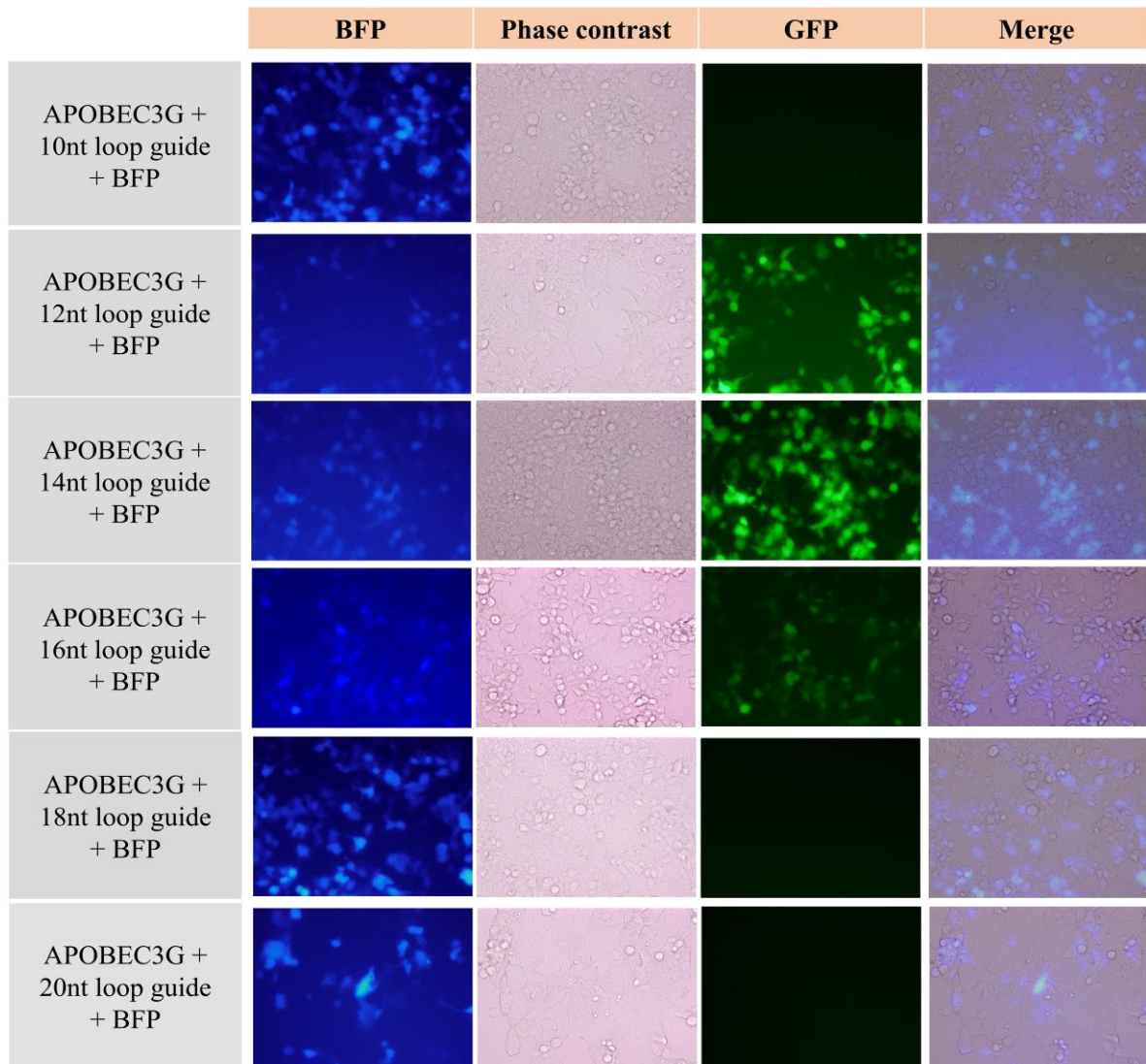
**Fig 5**: Fluorescence micrographs of HEK293T were obtained at 48h after transfection. The images on the left panel are BFP vision, the images on the second left panel are phase contrast, the images on the third left panel are GFP vision, and the right panel is merged.

From the figure 5, we found that the fluorescent signal can be detected in the 12, 14 and 16nt loop guide. We can't detect any signal in the 10, 18 and 20nt loop guide.

In the previous study, they introduced D317W into the APOBEC3G, and the D317W mutation showed higher base preference to 5'- TC [24]. We constructed the D317W APOBEC3G mutation type and co-transfected with guide RNA plasmids and BFP plasmid into HEK293T cells. The observation was performed by Keyence BZ-8000 fluorescence microscopy.

	BFP	Phase contrast	GFP	Merge
A3G D317W + 10nt loop guide + BFP				
A3G D317W+ 12nt loop guide + BFP				
A3G D317W+ 14nt loop guide + BFP				
A3G D317W+ 16nt loop guide + BFP				
A3G D317W+ 18nt loop guide + BFP				
A3G D317W+ 20nt loop guide + BFP				

Fig 6: Fluorescence micrographs of HEK293T were obtained at 48h after transfection. The images on the left panel are BFP vision, the images on the second left

panel are phase contrast, the images on the third left panel are GFP vision, and the right

panel is merged.

From the figure 5, we found the same result as wild type APOBEC3G. The

fluorescent signal can be detected in the 12, 14 and 16nt loop guide, but can't detect any

signal in the 10, 18 and 20nt loop guide.

We extracted the total RNA from the transfected cells for confirming the editing

efficiency of all conditions (**Fig 7**). The bar graph showed that when the guide RNA to

induce 14nt loop, APOBEC3G and D317W variant showed the highest editing efficiency

in all conditions. The editing efficiency of APOBEC3G D317W variant was higher than

the natural APOBEC3G. In the figure 8 and 9, same as the result of APOBEC3A, we

didn't detect any deamination signal of other C except the target C on the nucleotide loop

from the sanger sequencing result.
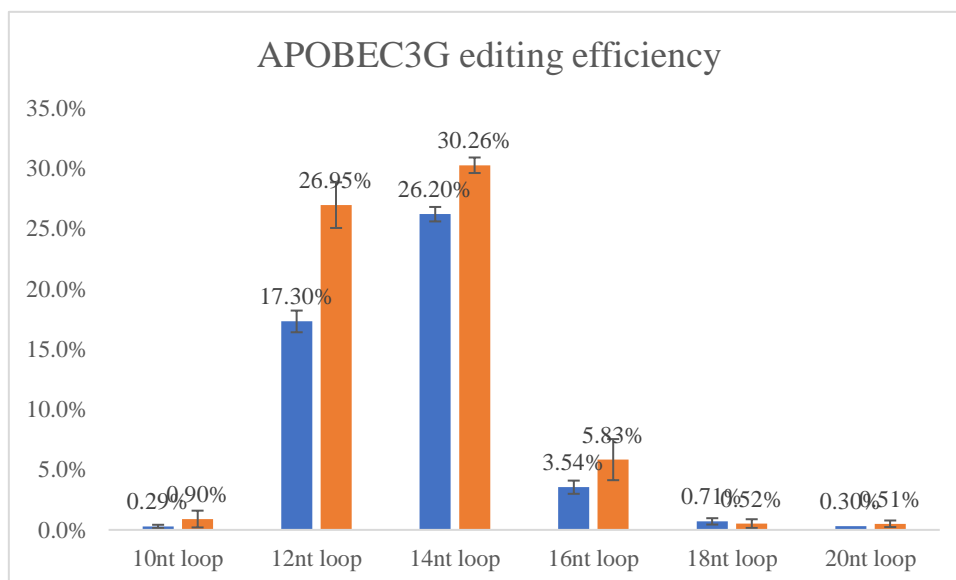
Fig 7: Bar graph showing the editing efficiency (%) calculated from the sequencing results using the peak height ratio method. Mean and standard error of the mean (SEM) (n=3).
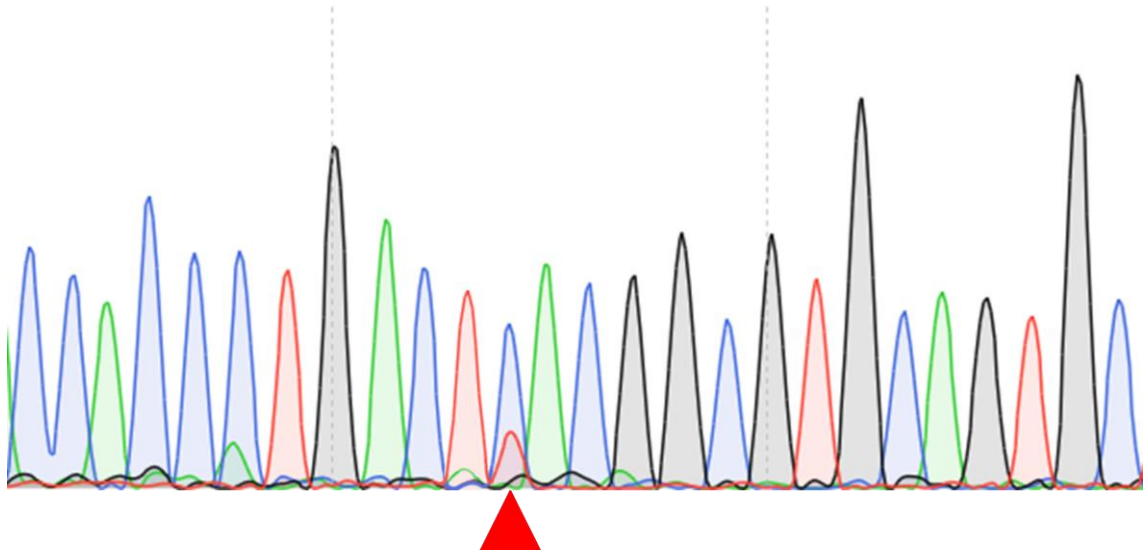
Fig8: Sanger sequencing result of wild type APOBEC3G with 14nt loop guide RNA
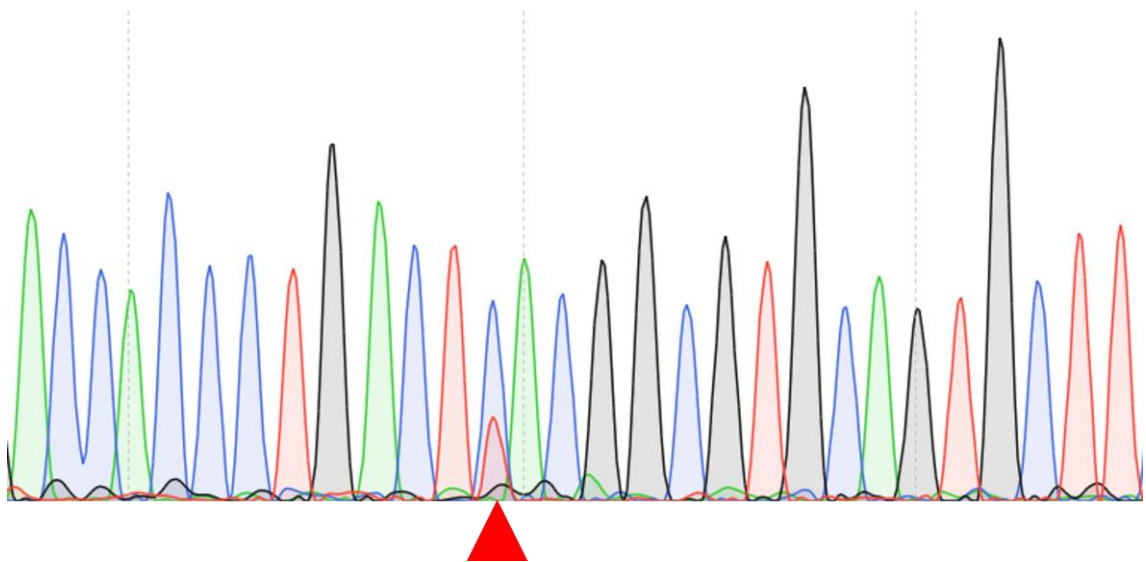
Fig9: Sanger sequencing result of APOBEC3G D317W mutation with 14nt loop guide RNA

4.3.3 Examination of the guide RNA on artificial APOBEC1 editing efficiency

After we observed the result of APOBEC3A and APOBEC3G, we assumed whether using the APOBEC1 system in the previous study, combined with this new guide RNA, could improve the editing efficiency. We co-transfected origin guide RNA or three types of guide RNA with the artificial APOBEC1 RNA editing system [23]. The observation was performed by Keyence BZ-8000 fluorescence microscopy.
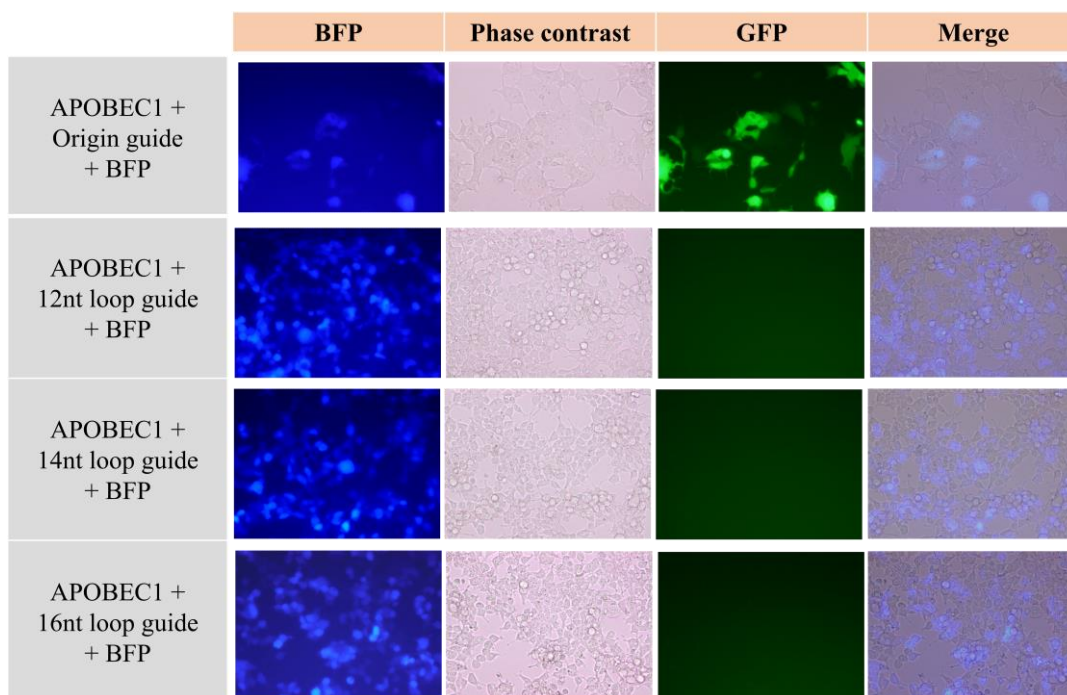
Fig 10: Fluorescence micrographs of HEK293T were obtained at 48h after transfection. The images on the left panel are BFP vision, the images on the second left panel are phase contrast, the images on the third left panel are GFP vision, and the right panel is merged.
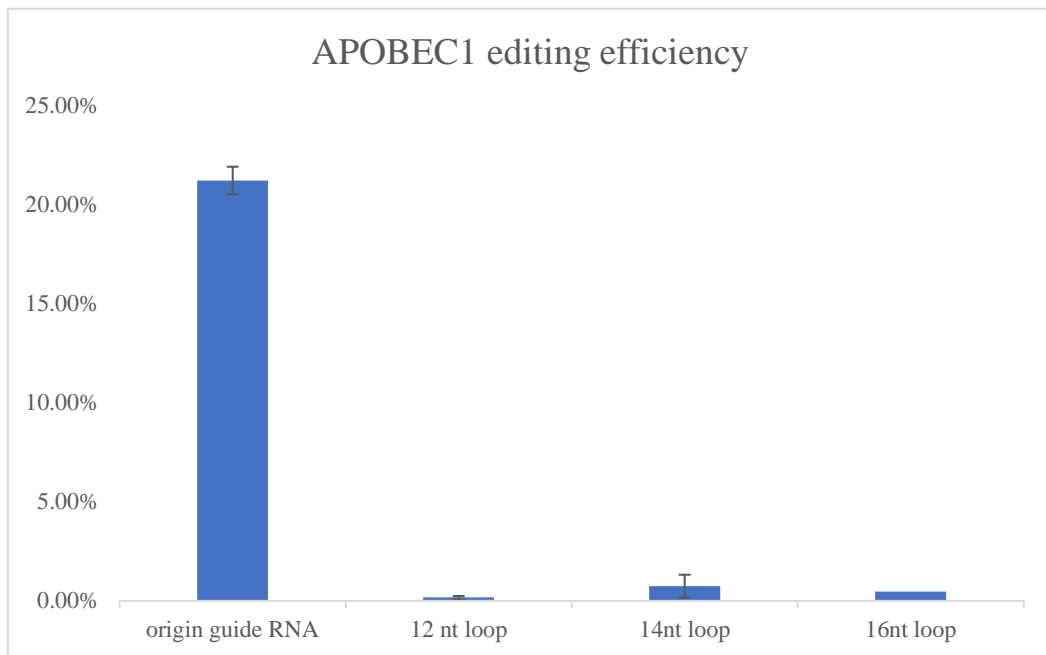
Fig 11: Bar graph showing the editing efficiency (%) calculated from the sequencing results using the peak height ratio method. Mean and standard error of the mean (SEM) (n=3).
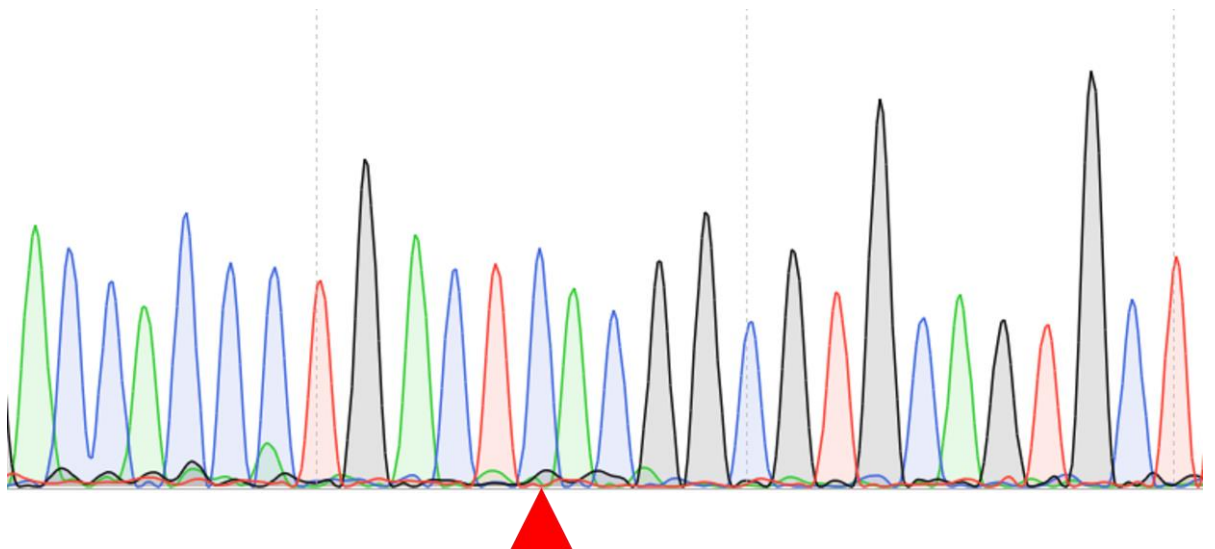
Fig 12: Sanger sequencing result of APOBEC1 with 14nt loop guide RNA

From the figure 10 and 11, we found that even the artificial APOBEC1 RNA editing system was co-transfected with all types of guide RNA, we can't detect any fluorescent signal or editing except the cells had transfected by origin guide RNA. In the sanger sequencing result, different from the APOBEC3A and APOBEC3G, not only the target C, but also other C didn't show any deamination result (Fig 12).

We also transfected the original guide RNA with each type of fusion protein into HEK 293T cells. But we can't detect any fluorescent signal from the APOBEC3A, wild type APOBEC3G and APOBEC3G D317W condition group (Fig 13)
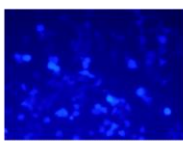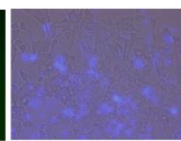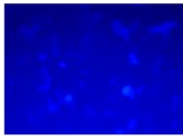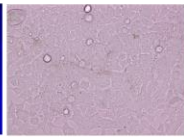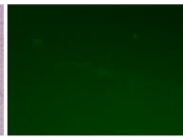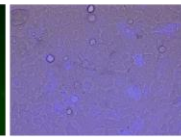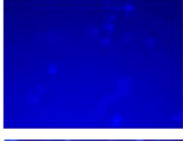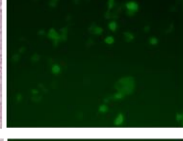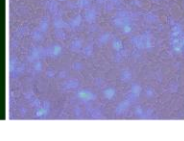
BFP	Enzyme	Guide RNA	BFP	Phase contrast	GFP	Merge
+	APOBEC3A	+				
+	APOBEC3G	+				
+	APOBEC3G D317W	+				
+	APOBEC1	+				

Fig 13: Fluorescent images of original guide RNA transfected with each type of fusion protein

4.3.4 Confirmation of fusion protein mRNA expression level

Due to the transfection efficiencies of different plasmids, we confirmed the expression of fusion protein mRNA by qPCR. We set thee expression level of the MS2-APOBEC3A fusion protein mRNA as 1, then the relative expression level of the MS2-APOBEC3G, MS2-D317W APOBEC3G and MS2-APOBEC1 was 1.54, 0.76 and 1.35, respectively (Fig 13).
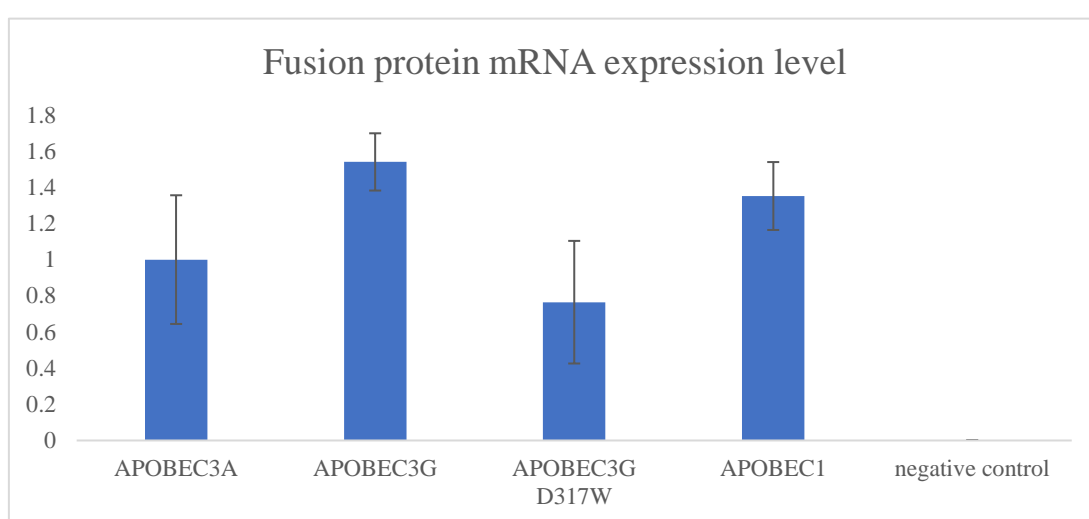
Fig 14: Bar graph showing the expression level of each type of fusion protein mRNA. The expression level of the APOBEC3A fusion protein mRNA was set as 1. Results are mean $\pm$ SEM (n=3).

Then, we calculated the expression level of each type of fusion protein mRNA and normalized the expression level of fusion protein mRNA to obtain the theoretical value of editing efficiency (Fig 14). The theoretical value of editing efficiencies were 41.00% for MS2-APOBEC3A, 17.01% for MS2-APOBEC3G, 39.58% for MS2-D317W APOBEC3G, 0.52% for MS2-APOBEC1. The calculated editing efficiency of

APOBEC3G D317W showed almost same level to APOBEC3A and stronger than the wild type APOBEC3G.
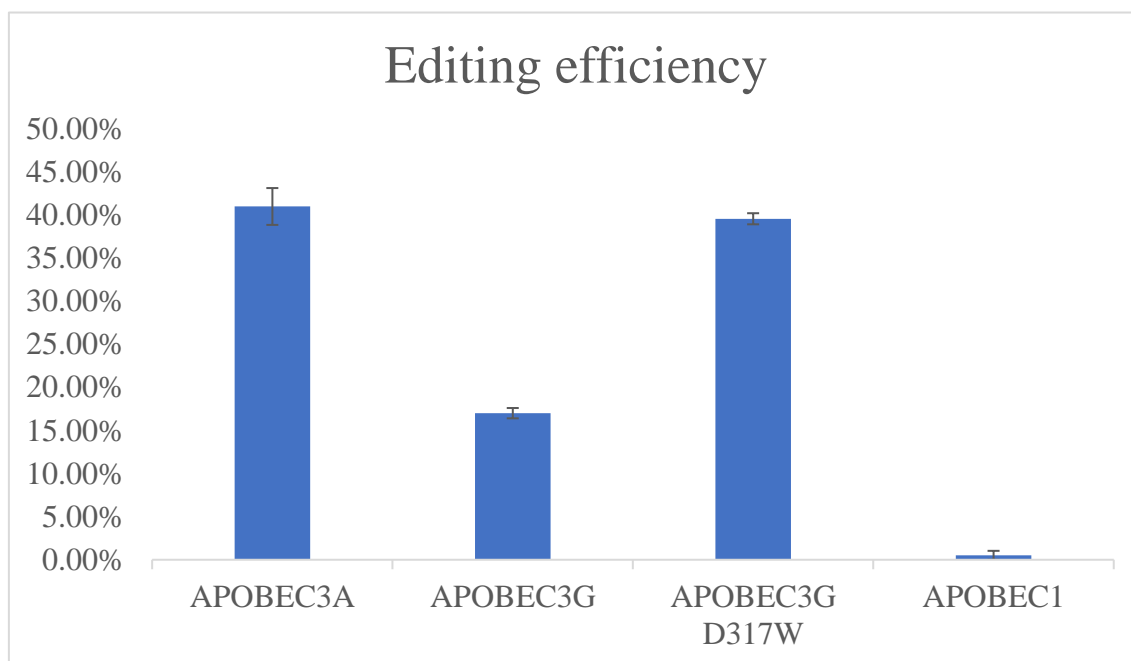
Fig 15: The theoretical value of editing efficiency of each fusion protein editing. Results are

mean ± SEM (n=3).

4.4 Discussion

In this study, we have utilized APOBEC3A and APOBEC3G to create the MS2-APOBEC3A/3G RNA editing system. Previous studies have demonstrated that APOBEC3A and APOBEC3G are capable of inducing C-to-U deamination on single-stranded RNA [21, 22]. The CURE system has also shown that APOBEC3A can be utilized for specific RNA deamination. In contrast to the RESCUE-S (ADAR2) system,

APOBEC3A and APOBEC3G do not induce A-to-I deamination on single-stranded RNA, which is advantageous for C-to-U RNA editing.

In Figure 13, no fluorescent signal was detected from the results of the original guide RNA transfected with APOBEC3A, APOBEC3G, or APOBEC3G D317W fusion protein into HEK293T cells. We hypothesized that APOBEC3A and APOBEC3G may not be flexible enough to accommodate the double-stranded RNA structure in their catalytic site.

As previously reported, APOBEC3A and APOBEC3G have preferences for substrate structure and nucleotide sequence, and the development of loop guide RNA has shown the potential for utilizing APOBEC3A and APOBEC3G for RNA editing [13]. The guide RNA can induce a loop structure on the target RNA, and the loop structure can affect the site-specific RNA editing by APOBEC3A and APOBEC3G [27]. In our results, APOBEC3A induced deamination editing in all types of loop guide RNA, but APOBEC3G only showed editing activity in three types of RNA editing. The target sequence we used and the formed loop structure affected the editing efficiency to some extent, as APOBEC3A and APOBEC3G have a strong preference for the base sequence and RNA stem-loop structure. However, when the guide RNA and the target RNA are paired by Watson-Crick pairing, the RNA secondary structure of the stem-loop formed is unpredictable. If this structure is not recognized and bound by APOBEC3A or APOBEC3G, it is difficult to catalyze the deamination reaction. It is possible that the catalytic site of APOBEC3A and APOBEC3G is not flexible enough to accommodate larger RNA loops, but we cannot confirm this hypothesis at this time. We will not discuss this possibility further in this paper due to the lack of evidence.

In RNA editing induced by APOBEC3A and APOBEC3G, the target sequence plays an important role. For APOBEC3A, C-to-U editing sites are commonly present within a CCAUCG sequence motif, with CAUC and its CCUC, CUUC, and UAUC (edited site underlined) [21]. Similarly, a previous study suggested [CGU]N[CU]C[AG] as a sequence motif commonly targeted by APOBEC3G (The residues within brackets are possibilities for a position, and the edited C is underlined) [22]. However, these nucleotide preferences may induce off-target editing on the loop sequence. Therefore, if the preference sequence does not exist in the target sequence, the editing efficiency will be decreased, or the deamination catalyzed will not be induced.

To improve RNA editing efficiency, amino acid substitution is another method that can be used to alter the deaminase function. In this study, we compared the editing efficiency of APOBEC3G and its D317W mutation. Although the D317W mutation slightly improves editing efficiency compared to the natural type, the improvement is not significant. This is because the D317W mutation only changes base sequence preference and cannot change RNA loop structure preference. We also found that we could not detect any deamination except the target C on the nucleotide loop (Fig 9). In a previous study, the human APOBEC3A mutation improved the specificity of single-strand RNA editing and reduced the editing activity on single-stranded DNA [29]. In another experiment, my colleague performed C-to-U RNA editing using artificial APOBEC1 deaminase with the original guide RNA [23]. Although we used the MS2-APOBEC1 system to perform transfection with loop guide RNA, we did not observe any restoration results under these conditions (Fig 11). We suggest that APOBEC1 may be better suited to the structure of double-stranded RNA than to single-stranded RNA.

We also used qPCR to confirm the mRNA expression levels of the fusion proteins and normalized the editing efficiency of each fusion protein accordingly. Although the D317W mutation in APOBEC3G showed a slight improvement in editing efficiency (approximately 46%), the mRNA expression level cannot replace the expression of the fusion protein. The calculated editing efficiency of APOBEC3G D317W suggests that it could be more efficient than the wild type in catalyzing reactions. However, in actual experiments, many factors can affect the editing efficiency, and it is challenging to achieve the theoretical editing efficiency.

In addition, the optimization of the loop guide RNA is a complex process that must be tailored to different RNA editing systems. In this study, we attempted to optimize the loop guide RNA to improve editing efficiency. However, due to the complexity of the guide RNA optimization process, we were unable to fully optimize the guide RNA in this project.

4.5 Conclusion

RNA editing is a crucial mechanism for regulating gene expression and expanding genetic diversity. Our research offers a novel approach to achieve site-specific C-to-U RNA editing using APOBEC3A/3G and the MS2 system. Our findings demonstrate the high adaptability of the MS2 system to RNA editing and the potential for forming protein complexes with other RNA editing enzymes to achieve site-specific RNA editing. These results suggest that there is still significant potential for optimizing the MS2 system for RNA editing and other applications. Thus, this study mainly focused on exploring the RNA editing induced by APOBEC3A and APOBEC3G.

4.6 Reference

[1]   Tang, W., Fei, Y. & Page, M. Biological significance of RNA editing in cells. Mol. Biotechnol. 52, 91–100 (2012).

[2]   Rees HA, Liu DR. Base editing: precision chemistry on the genome and transcriptome of living cells. Nat Rev Genet 19: 770 (2018)

[3]   Mao S, Liu Y, Huang S, Huang X, Chi T. Site-directed RNA editing (SDRE): Off-target effects and their countermeasures. J Genet Genomics 46: 531 – 535. (2019)

[4]   Montiel-Gonzalez MF, Diaz Quiroz JF, Rosenthal JJC. Current strategies for Site-Directed RNA Editing using ADARs. Methods San Diego Calif 156: 16 – 24 (2019)

[5]   Vogel P, Stafforst T. Critical review on engineering deaminases for sitedirected RNA editing. Curr Opin Biotechnol 55: 74 – 80 (2019).

[6]   Reardon S. Step aside CRISPR, RNA editing is taking off. Nature 578: 24 – 27 (2020)

[7]   Vogel, P.; Moschref, M.; Li, Q.; Merkle, T.; Selvasaravanan, K.D.; Li, J.B.; Stafforst, T. Efficient and Precise Editing of Endogenous Transcripts with SNAP-Tagged ADARs. Nat. Methods 2018, 15, 535–538.

[8]   Vallecillo-Viejo, I.C.; Liscovitch-Brauer, N.; Montiel-Gonzalez, M.F.; Eisenberg, E.; Rosenthal, J.J. Abundant Off-Target Edits from Site-Directed RNA Editing Can Be Reduced by Nuclear Localization of the Editing Enzyme. RNA Biol. 2018, 15, 104–114.

[9]   Cox, D.B.; Gootenberg, J.S.; Abudayyeh, O.O.; Franklin, B.; Kellner, M.J.; Joung, J.; Zhang, F. RNA Editing with CRISPR-Cas13. Science 2017, 358, 1019–1027.

[10]   Azad, M.T.; Bhakta, S.; Tsukahara, T. Site-Directed RNA Editing by Adenosine Deaminase Acting on RNA for Correction of the Genetic Code in Gene Therapy. Gene Ther. 2017, 24, 779–786.

[11]   Komor A. C., Kim Y. B., Packer M. S., Zuris J. A., Liu D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 533 (7603), 420–424.

[12]   Abudayyeh, O.O.; Gootenberg, J.S.; Franklin, B.; Koob, J.; Kellner, M.J.; Ladha, A.; Joung, J.; Kirchgatterer, P.; Cox, D.B.; Zhang, F. A Cytosine Deaminase for Programmable Single-Base RNA Editing. Science 2019, 365, 382–386.

[13]   Huang X, Lv J, Li Y, Mao S, Li Z, Jing Z, Sun Y, Zhang X, Shen S, Wang X, Di M, Ge J, Huang X, Zuo E, Chi T. Programmable C-to-U RNA editing using the human APOBEC3A deaminase. EMBO J. 2020 Nov 16;39(22):e104741. doi: 10.15252/embj.2020104741. Epub 2020 Oct 15. Erratum in: EMBO J. 2021 May 3;40(9):e108209.

[14]   Chen SH, Habib G, Yang CY, Gu ZW, Lee BR, Weng SA, Silberman SR, Cai SJ, Deslypere JP, Rosseneu M et al. 1987. Apolipoprotein B-48 is the product of a messenger RNA with an organ-specific in-frame stop codon. Science 238: 363-366.

[15]   Hospattankar AV, Higuchi K, Law SW, Meglin N, Brewer HB, Jr. 1987. Identification of a novel in-frame translational stop codon in human intestine apoB mRNA. Biochem Biophys Res Commun 148: 279-285.

[16]   Powell LM, Wallis SC, Pease RJ, Edwards YH, Knott TJ, Scott J. 1987. A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine. Cell 50: 831-8

[17]   Bransteitter, R., Prochnow, C. & Chen, X. S. The current structural and functional understanding of APOBEC deaminases. Cell. Mol. Life Sci. 66, 3137–3147 (2009).

[18]   Cullen BR. 2006. Role and mechanism of action of the APOBEC3 family of antiretroviral resistance factors. Journal of Virology 80:1067–1076.

[19]   Chiu YL, Greene WC. 2008. The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. Annual Review of Immunology 26:317–353.

[20]   Harris RS, Dudley JP. 2015. APOBECs and virus restriction. Virology 479-480:131–145.

[21]   Sharma S, Patnaik SK, Taggart RT, Kannisto ED, Enriquez SM, Gollnick P, Baysal BE. APOBEC3A cytidine deaminase induces RNA editing in monocytes and macrophages. Nat Commun. 2015 Apr.

[22]   Sharma S, Patnaik SK, Taggart RT, Baysal BE. The double-domain cytidine deaminase APOBEC3G is a cellular site-specific RNA editing enzyme. Sci Rep. 2016 Dec

[23]   Bhakta, S., Sakari, M. & Tsukahara, T. RNA editing of BFP, a point mutant of GFP, using artificial APOBEC1 deaminase to restore the genetic code. Sci Rep 10, 17304 (2020).

[24]   Rathore A, Carpenter MA, Demir Ö, Ikeda T, Li M, Shaban NM, Law EK, Anokhin D, Brown WL, Amaro RE, Harris RS. The local dinucleotide preference of APOBEC3G can be altered from 5'-CC to 5'-TC by a single amino acid substitution. J Mol Biol. 2013 Nov 15;425(22):4442-54.

[25]   Heim R, Prasher DC, Tsien RY. Wavelength mutations and posttranslational autoxidation of green fluorescent protein. Proc Natl Acad Sci. 1994;91:12501–4.

[26]   Coelho, M.A., Li, S., Pane, L.S. *et al.* BE-FLARE: a fluorescent reporter of base editing activity reveals editing characteristics of APOBEC3A and APOBEC3B. *BMC Biol* **16**, 150 (2018).

[27]   Sharma S, Baysal BE. 2017. Stem-loop structure preference for site-specific RNA editing by APOBEC3A and APOBEC3G. PeerJ 5:e4136

[28]   Tetsuto Tohama, Matomo Sakari and Toshifumi Tsukahara. Development of a Single Construct System for Site-Directed RNA Editing Using MS2-ADAR. International Journal of Molecular Sciences 2020; 21(14): 4943.

[29]   Tang G, Xie B, Hong X, Qin H, Wang J, Huang H, Hao P, Li X. Creating RNA Specific C-to-U Editase from APOBEC3A by Separation of Its Activities on DNA and RNA Substrates. ACS Synth Biol. 2021 May 21;10(5):1106-1115.

Chapter5: General discussion

5.1 Discussion

After the discovery of the CRISPR-Cas system, genome editing has garnered significant attention. Despite the availability of other reprogrammable genome editing tools such as TALEN and zinc finger nucleases, they are primarily used to create genetically modified organisms (GMOs) for research. The development of gene editing tools heavily relies on algorithms for designing the guide sequence, preventing off-target effects, and determining target specificity. The efficacy, specificity, and safety of the editing process are crucial considerations when using genome editing techniques. The CRISPR-Cas system requires a homology-mediated repair system to repair the cut and inserted DNA, which is only present in dividing cells, making it challenging to use in cells like nerve and muscle cells that do not divide. The potential of modifying other genes when using CRISPR on germ cells or an embryo is a concern, potentially affecting the functionality of multiple organs. Despite this, CRISPR gene editing has the potential to cure over 6,000 identified genetic illnesses since its launch in 2012. Both China and the USA have made genuine applications of CRISPR to treat humans, including cancer and blood problems. Although the technology is still being scrutinized, CRISPR is more precise than conventional gene therapy, making it a promising option for treating some disorders that gene therapy has not been able to address. The reputation of CRISPR was damaged in 2018 when a Chinese researcher altered a gene in embryos, leading to the birth of two girls and sparking debate over the technology's ethical implications. However, current CRISPR trials do not pose the same ethical challenges as they are being tested on

adults and children and do not cause inherited DNA modifications. The use of CRISPR and deaminases to correct point mutations in DNA has opened up a new frontier in DNA correction.

One significant post-transcriptional change to the genetic information contained in a living organism's genome is RNA editing, which is facilitated by deaminases from several families that can perform A-to-I, C-to-U, and even U-to-C type RNA editing. Through the use of targeted mRNA, artificial RNA editing allows for the fine-tuning of protein activity without having to modify the intricate and highly ordered genomic DNA. The deaminases naturally bind to their targets in a sequence-specific manner using related structures. For example, APOBEC targets the cytidine next to "mooring sequences," while ADARs bind to the dsRNA created by the inverted Alu repeats in RNA. However, these deaminases can be modified to target a specific nucleotide at any position. This natural deaminase can be harnessed and engineered, offering a great opportunity for therapeutic use in the future to cure genetic illnesses caused by point mutations. A-to-I editing can alter 12 out of 20 amino acids. The technique that I developed can successfully target any cytosolic RNA, making it suitable for use at any stage of development.

In Chapter 2, we demonstrated how the well-known MS2 mechanism can be utilized to direct the ADAR1 deaminase domain to achieve site-specific A-to-I RNA editing. To improve the editing efficiency, we replaced the 6X MS2 stem-loop RNA with a 12X MS2 stem-loop RNA. However, we were unable to detect any fluorescent signal in all transfection conditions due to the distance between the antisense and stem-loop regions. This finding highlights the importance of the distance between the antisense and stem-loop regions for effective RNA editing.

In Chapter 3, we aimed to further improve the efficiency of RNA editing based on previous research. My colleague had developed the 1-1 stem-loop guide RNA and achieved an editing efficiency of approximately 65%. Building upon his work, we attempted to increase the number of stem-loops and incorporate mismatch bases. Interestingly, our results differed from those of previous studies. We found that when the number of stem-loops was the same at two sites of the antisense RNA, the editing efficiency was higher compared to other conditions. Moreover, when the mismatch base was U, we observed higher editing efficiency than with other bases, which contradicts the findings of a previous study [5]. This indicates that the guide RNA must be optimized for different RNA editing systems to achieve better efficiency.

In Chapter 4, we utilized the MS2 system to harness APOBEC3A and APOBEC3G for performing C-to-U RNA editing. We also developed a loop guide RNA for RNA editing, which differs from the base mismatch guide RNA, as it induces a loop structure on the target RNA by using a single-strand RNA. The loop can then be identified by the single-strand catalytic enzyme, such as APOBEC3A or APOBEC3G. This novel method enables the catalytic enzyme, which previously could only catalyze single-strand RNA, to catalyze double-strand RNA editing. Additionally, we investigated whether the length of the loop affects the editing efficiency and found that it does indeed affect the efficiency of the editing process. Furthermore, we discovered that mutations in the catalytic enzyme also affect the editing efficiency. These findings indicate that further optimization of the APOBEC family system and loop guide RNA is necessary before this technique can be applied in disease treatment in the future.

The recent advances in gene editing are very exciting. The CRISPR-Cas13 system has been shown to perform successful A-to-I and C-to-U RNA editing in previous studies.

Additionally, the size of CRISPR has recently become smaller [6], indicating that it is constantly being updated. This guide system can also be applied to other deaminases. However, research on RNA editing using the MS2 system is limited, and there are few advanced studies on MS2.

In conclusion, the recent advancements in RNA editing, particularly in the MS2-ADAR and MS2-APOBEC systems, have expanded the research on RNA editing related to MS2. While the improvement of MS2-ADAR did not significantly increase the editing efficiency, it highlighted the possibility of further optimization for this system. The development of the loop guide RNA and the APOBEC3 family have shown great potential for the MS2 system in C-to-U RNA editing. These developments are important landmarks in the correction of point mutations for therapeutic purposes in the near future.

5.2 Reference

[1]   Kim D, Bae S, Park J, Kim E, Kim S, Yu HR, Hwang J, Kim JI, Kim JS. Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. Nat Methods. 2015 Mar;12(3):237-43, 1 p following 243.

[2]   Matsoukas IG. Commentary: Programmable base editing of A·T to G·C in genomic DNA without DNA cleavage. Front Genet. 2018 Feb 7;9:21.

[3]   Cyranoski D. CRISPR gene-editing tested in a person for the first time. Nature. 2016 Nov 24;539(7630):479.

[4]   Reardon, S. First CRISPR clinical trial gets green light from US panel. *Nature* (2016).

[5]   Qu L, Yi Z, Zhu S, Wang C, Cao Z, Zhou Z, Yuan P, Yu Y, Tian F, Liu Z, Bao Y, Zhao Y, Wei W. Programmable RNA editing by recruiting endogenous ADAR using engineered RNAs. Nat Biotechnol. 2019 Sep;37(9):1059-1069.

[6]   Gupta R, Ghosh A, Chakravarti R, Singh R, Ravichandiran V, Swarnakar S, Ghosh D. Cas13d: A New Molecular Scissor for Transcriptome Engineering. Front Cell Dev Biol. 2022 Mar 31;10:866800.

List of Publications

1. Jiarui Li, Tomoko Oonishi, Guangyao Fan, Matomo Sakari, Toshifumi Tsukahara. Increasing the editing efficiency of the MS2-ADAR system for site-directed RNA editing. Applied Sciences. 2023, 13(4), 2383

2. Jiarui Li, Guangyao Fan, Matomo Sakari, Toshifumi Tsukahara. Programmable C-to-U RNA editing using human APOBEC3A and APOBEC3G deaminase. (Ready to summit)

ACKNOWLEDGE

I would like to express my sincere gratitude to my esteemed supervisor, Professor Dr. Toshifumi Tsukahara, for his unwavering support, patience, and guidance throughout my doctoral studies. His invaluable advice and encouragement have been instrumental in broadening my research scope and allowing me to work freely in various fields. I owe my accomplishments during my Ph.D. study to his cooperation and wisdom.

I am also deeply grateful to Associate Professor Dr. Takumi Yamaguchi for his invaluable guidance as my second research supervisor. His suggestions and discussions have been immensely helpful in completing my thesis.

I would like to extend my heartfelt appreciation to my minor research supervisor, Professor Dr. Takahiro Hohsaka, for giving me the opportunity to carry out minor research experiments in his laboratory and for his valuable guidance, without which this study would not have been possible.

I am also thankful to Senior Lecturer Dr. Takayoshi Watanabe for the invaluable experience of learning new techniques under his guidance.

I express my sincere gratitude to Dr. Matomo Sakari for his constant support and motivation through his valuable discussions on experimental methods and data analysis.

I would also like to thank Peiyi Chen, Ruchika Mishra, John Munene, Tetsuto Tohama, Wenhao Jin, Mitsuki Furuya, Junling Mo, Heeraman, Tomoko Onishi, Zhaoyuan Dai, Wakana Chinen, Chisato Okhudaira, Yuki Nakagawa, Ryutaro Shimo, Tomokazu Suiyama, Fumitaka Nakano, and Shosei Kanako for their constant support and motivation during my entire PhD journey at JAIST.