Artificial deaminase system for restoration of genetic code

Sonali Bhakta

Japan Advanced Institute of Science and Technology

Doctoral Dissertation

Artificial deaminase system for restoration of genetic code

Sonali Bhakta

Supervisor: Professor Dr. Toshifumi Tsukahara

Graduate School of Advanced Science and Technology

Japan Advanced Institute of Science and Technology

Materials Science

September, 2020

Doctoral Dissertation

Artificial deaminase system for restoration of genetic code

Main theme Advisor	Professor Toshifumi Tsukahara
Sub-theme Advisor	Professor Kenjo Fuzimoto

Judging Committee

Head of Committee	Professor Dr. Toshifumi Tsukahara
Internal Member	Professor Dr. Kenzo Fujimoto
Internal Member	Professor Dr. Takahiro Hosaka
Internal Member	Associate Professor Dr. Hidekazu Tsutsui
External Member	Professor Dr. Yoshitsugo Aoki

SONALI BHAKTA

s1720427

Graduate School of Advanced Science and Technology

Japan Advanced Institute of Science and Technology

Materials Science

September, 2020

Table of Contents

Table of Contents

Table of Contents

Table of Contents

Table of Contents

ABSTRACT

Site directed mutagenesis is an exceptionally viable way to deal with recode genetic information. Legitimate connecting of the synergist area of the RNA altering catalytic deaminase Adenosine Deaminase Acting on RNA (ADAR) or Cytidine Deaminase Acting on RNA (APOBEC) to an antisense direct RNA can change over explicit adenosines (As) to inosines (Is), with the last perceived as guanosines (Gs) during the translation procedure or Cytidines (Cs) to Uridines (Us). In this study, endeavors have been made to engineer the deaminase domain of ADAR1 and MS2 framework to target explicit A residues to reestablish G→A transformations. The target mRNA comprised of an ochre (TAA) stop codon, created from the TGG codon encoding amino acid 58 (Trp) of improved green fluorescent protein (EGFP). This framework had the capacity to change over the stop codon (TAA) to a decipherable codon (TGG), accordingly reestablishing fluorescence in a cell framework, as appeared by JuLi fluorescence and LSM confocal microscopy. The specificity of the editing was affirmed by Polymerase Chain Reaction-Restriction Fragment Length Polymorphism (PCR-RFLP), as the restored GFP mRNA could be cleaved into fragments of 160 and 100 base pairs, the absolute amplified length was 260 bp. Sanger's sequencing illustration with both the sense and antisense primers indicated that the reclamation rate was higher for the 5'A than for the 3'A. This system might be very useful for treating genetic diseases that result from the G to A point mutations.

Further an artificial editase of RNA was engineered by combining the deaminase domain of APOBEC1 (apolipoprotein B mRNA editing catalytic polypeptide 1) with a guideRNA (gRNA) which is complementary to target mRNA. In this artificial enzyme system, gRNA is bound to MS2 stem-loop, and deaminase domain, which has the ability to convert mutated target nucleotide C-to-U, is fused to MS2 coat protein. As a target RNA, here RNA encoding Blue Fluorescent Protein (BFP) was used which is derivative of the gene encoding GFP by 199T>C mutation. Upon transient expression of both components (deaminase and gRNA), GFP fluorescence was observed by confocal microscopy, indicating that mutated 199th C in BFP had been converted to U, restoring original sequence of GFP. This result was confirmed by PCR-RFLP and Sanger's sequencing using cDNA from transfected cells, revealing an editing efficiency of approximately 21%. Deep RNA sequencing result showed that off-target editing was sufficiently low in this system.

Later on, improving U6 promoter activity by CMV enhancer or promoter in target cells have been demonstrated to be a viable method to obtain satisfactory percentage of editing efficiency. The placement of a CMV enhancer nearby to U6 promoter or hybrid CMV-H1 promoter has been accounted for improving the efficiency of RNAi or shRNA delivery *in vivo*. From the experimental data it has been found that in case of the CMV promoter controlled process of RNA editing where both the deaminase and guideRNA constructs were prepared under the control of the pol II CMV promoter, the editing efficiency was lesser comparing to the U6 promoter containing guideRNA or in single construct having combined approach of CMV in deaminase domain and U6 promoter in guideRNA construct. From the PCR-RFLP (band intensity) data had also been observed that with the increase of the concentration of the deaminase or the guideRNA the restoration percentage had also increased. The editing efficiency has been calculated from the peak height of the Sanger's sequencing data. After the calculation of the efficiency it was found that in case of the CMV controlled approach the rate was 21.02% whereas in case of the U6 controlled and in case of single construct the restoration rate was 39.37% and 41.65%, respectively.

For performing the *in vivo* application of the developed artificial enzyme system the macular mouse model was chosen. The mutation in the P type copper transporting ATPase (ATP7A) gene is responsible for the Menkes kinky hair disease, where T-to-C mutation happens. It was found from our data that all the heterozygous female (Ml/+), normal littermate male (+/y) and hemizygous male (Ml/y) had increased the body weight as usual up to 10 days of age. After that the body weight of heterozygous female (Ml/+) and normal littermate (+/y) increased significantly at 14 days as well but in case of the hemizygous male (Ml/y), its body weight significantly reduced at 14^{th} day of age. The peak area and peak height from the Sanger's sequencing analysis was measured by ImageJ (NIH) software. From the calculation it was found that by using the APOBEC1 deaminase and U6-21bp upstream-MS2-6X guideRNA 12.17% and 16.25% of the genetic code was restored in the macular mouse derived fibroblast cells by peak area and peak height, respectively. Where the deaminase and guideRNA, were two different constructs. After that single construct was applied where the deaminase was controlled by pol II CMV vector and guideRNA was under the control of pol III U6 promoter, in the same plasmid vector. The peak area and peak height from the Sanger's sequencing analysis were measured by using ImageJ (NIH) software. From the calculation we found that by using the APOBEC 1 deaminase and U6-MS2-6X-21bp upstream 27.20% and 26.09% of the genetic code was restored, respectively calculated from peak area and peak height. Afterwards, the 1X MS2 on either side of guide sequence containing guideRNA construct was introduced along with the APOBEC 1 deaminase. Similarly the sample was sequenced for observing the editing rate. Editing rate was calculated both by peak area and peak height. I found that editing rate was 36.66% and 34%, respectively by peak area and peak height. For any developed system it is more important that the application could be achieved for the purpose of treatment. The developed artificial deaminase system for both the A-to-I and C-to-U editing could be applied to the through the viral vector (AAVs) easily into the host body for the therapeutic purpose. The proper application of the developed artificial deaminase system for the treatment of the patients who are suffering from such type of mutagenic diseases could open a new era in the field of genetic diseases.

Key words: Genetic code, RNA editing, Deaminase domain, Macular mouse, ATP7A gene

List of Figures

List of Figures

List of Figures

List of Figures

List of Figures

List of Figures

List of Figures

List of Tables

CHAPTER I

General Introduction

INTRODUCTION

1.1 Genetic engineering:

Genetic engineering is also termed as genetic modification or manipulation, is the direct manipulation of an organism's genes using biotechnology. It consists of a set of technologies which are used for regulating functions of intracellular target gene or expression has been widely used in basic research, medicinal and applications as therapy (1). It is also utilized for changing the genetic makeup of cells, including the transfer of genes within and across species boundaries to produce improved or novel organisms. New DNA is obtained by either isolating or replicating the hereditary material of interest utilizing recombinant DNA strategies or by artificially incorporating or synthesizing the DNA. Development of construct is generally made and used to embed this DNA into the host living being. The primary motivation behind the genetic engineering methodologies is to control the capacity of intracellular proteins associated with the organic procedures of intrigue.

1.2 Genome editing

Genome alteration or genome designing is a kind of genetic engineering where DNA is embedded, erased, adjusted or supplanted in the genome of a living organism. As of late progressions in genome altering innovations have generously improved the capacity to roll out exact improvements in the genomes of eukaryotic cells. In recent times several technologies of genome editing have made it possible to alter target genomic information (1-5). Notwithstanding to programmable regulation, one of the most significant highlights of the genome editing is that it can provide a permanent alteration to the targeted gene in the specific cell, while this perpetual modification can

successfully limit the target protein expressions. Such techniques may present wellbeing dangers if a blunder happens (6).

1.3 RNA editing

RNA editing is another significant mechanism for directing hereditary versatility through the age of elective protein items from a solitary basic quality. Substitutional RNA editing utilizes an assortment of genetic mechanisms, the biochemical premise of which has been explained following the advancement of *in vitro* measures that reiterate vital components of this procedure. For the most part two sorts of substitutional RNA exist in mammals, namely A-to-I and C-to-U RNA editing (7, 8), whereas another type U-to-C substitution is found mainly in lower plants. Important biochemical differentiations between these two procedures provide an informative basis for understanding the mechanisms of C-to-U RNA editing and the adjustments control the target particularity.

1.4 Types of RNA editing:

There are several types of RNA editing events in living organisms. Among these A-to-I, C-to-U and U-to-C are the most common types. A-to-I and C-to-U RNA editing are generally caused by ADARs and APOBEC-AID deaminase family respectively. However, the enzymes responsible for U-to-C editing does not discover yet although it is the abundant phenomenon in lower plant species and rare in animals.

1.5 Advantages and disadvantages of the genome editing and RNA editing

The principle motto behind the genetic engineering techniques is to control the functional capacity of intracellular proteins engaged with the biological processes of intrigue. By recent times several genome editing technologies such as CRISPR-Cas system, TALEN, ZFN etc. have made it conceivable to control target genomic

information (10-14). In addition to programmable regulation, one of the most significant features of the genome editing is that it can provide a change to the targeted gene in cellular system, while this effect can control the expression or production of targeted protein efficaciously (15). However, genome editing has some drawbacks, as if any alteration at off target level occurs that will be permanent in case of the genome editing and will affect the genome sequence consequently, whereas mismatch will not be permanent in RNA editing; even not affect the genome sequence. RNA editing is much more for treatment purpose compared to genome editing. Particularly for very popular CRISPR-Cas genome editing system, homology-mediated repair system is needed to repair between cut and inserted DNA which is only present in dividing cells. In case of the non-dividing cells like nerve and muscle cells, this system facing challenges. In association with the technical challenges and bioethical issue, it's still controversial regarding the clinical application of this technology .Therefore, I have chosen the RNA editing as my choice of approach to build up an artificial enzymatic system for restoration of genetic information.

RNA sequences can be edited with regarding their genome-encoded arrangement like how an editorial manager changes letters or words in an original copy. The principal recognized instance of such RNA editing was the post-transcriptional addition or erasure of non-encoded uridine (U) residues into the mitochondrial RNA of trypanosomes. This technique generates accurate, 'readable' open reading frames that cannot be reasoned from sequences of genomic arrangements. In higher eukaryotes, base alteration is the significant kind of RNA editing, with the best-portrayed model being deamination responses, in which cytidine (C) or adenosine (A) is converted to uracil (U) and inosine (I), respectively. Moreover, a couple of U-to-C and G-to-A

alterations have been accounted for, which may conceivably result from trans-amination reactions (Figure 2). If such base modifications happen in coding areas of mRNA, the amino-acid particularity of codons can be changed, bringing about the blend of protein isoforms not predictable from genome sequences. Be that as it may, in the event that it is happening in the anticodons of tRNAs, base deamination can increment or even change the codon-recognition capacity during translation.

1.6 ADAR (Adenosine deaminases acting on RNA) enzyme:

Adenosine deaminases acting on RNA (ADARs) are the enzymes for RNA editing which have the capacity to execute the hydrolytic deamination of the adenosine (A) in the context of RNA polymers, at the corresponding nucleotide positions converting to inosine (I) (22, 23, 24). This reaction has a an assortment of practical ramifications for the RNA substrates including modulating thermal stability of base-paired structures, modifying the significance of codons in mRNAs (recoding), changing the splicing patterns of pre-mRNAs, altering miRNA targeting within 3 untranslated regions (UTRs), *etc*. (25-31). With the recognition of the large numbers of inosine (I) locales in the human transcriptome this circumstance is widespread (32). Also, mis-regulation of A-to-I RNA editing is associated with human ailments (33–38, 39). To be precise, alteration of one of the human genes responsible for A-to-I RNA editing (ADAR1) is a reason for the acquired immune disease such as Aicardi–Goutieres syndrome (40). Human cells express three distinct types of ADARs (adenosine deaminase acting on RNA), named ADAR1, 2, and 3, (Figure 3) which are possibly re-addressable for site-directed RNA editing (41, 42). ADAR1 and ADAR2 proteins have the well characterized activity of adenosine deamination (43). While ADAR3 is expressed in human brain however its capacity stays a mystery since no reactant action has been

accounted for the protein yet. The ADARs are particular in their structure with clearly identifiable RNA binding domains and also deaminase domains (Figure 1). In reality, RNA binding is largely dependent on double-stranded RNA binding domains (dsRBDs) present in each ADAR protein as multiple copies (44). dsRBDs tie up with any RNA duplex greater than ~16 bp, in a sequence independent manner (45). The existance of dsRBDs in the ADAR structure is predictable with the prerequisite for duplex RNA in ADAR substrates for effective deamination to take place (Figure 4). However, ongoing studies intended to guide editing to new positions with fusion proteins containing ADAR deaminase domains propose these domains likewise require duplex RNA for effective responses, even in absence of dsRBDs (46-48). While the RNA-binding properties of dsRBDs of ADARs have been all around well archived (49, 50), scarcely any reports have concentrated on the RNA restricting necessities of the ADAR synergist areas.

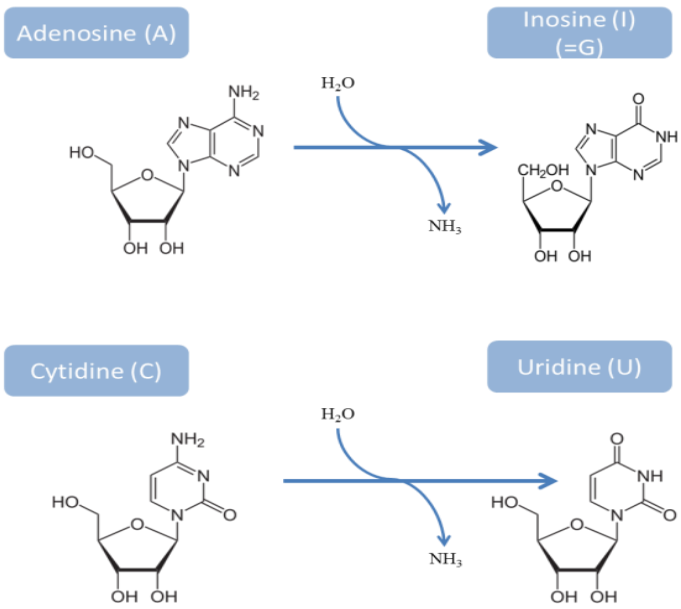
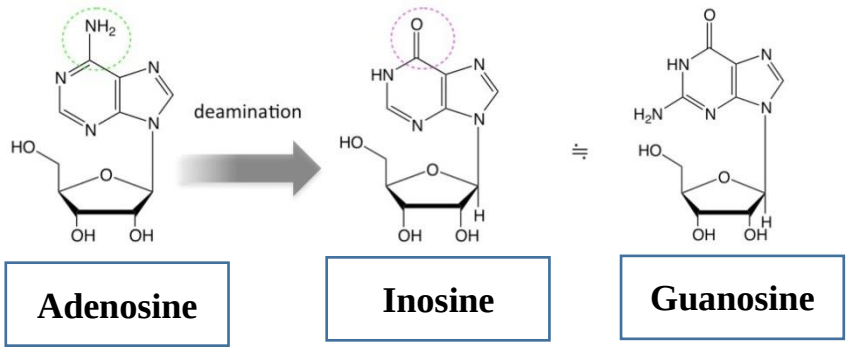
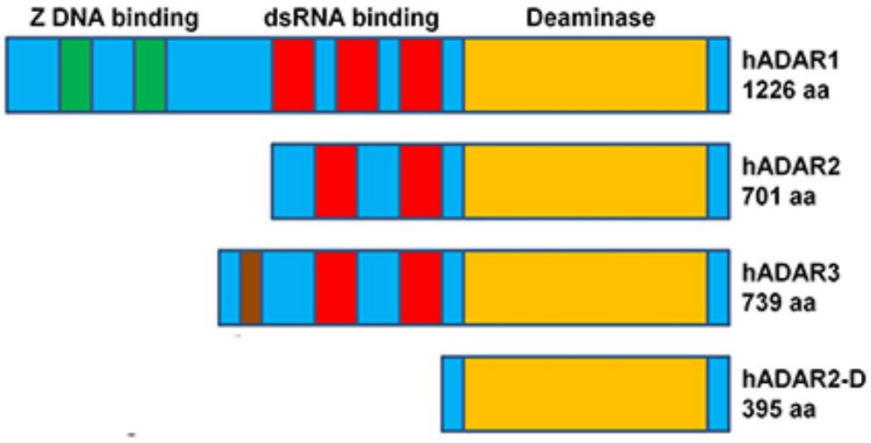


Fig 1. Chemical conversion of A-to-I and C-to-U



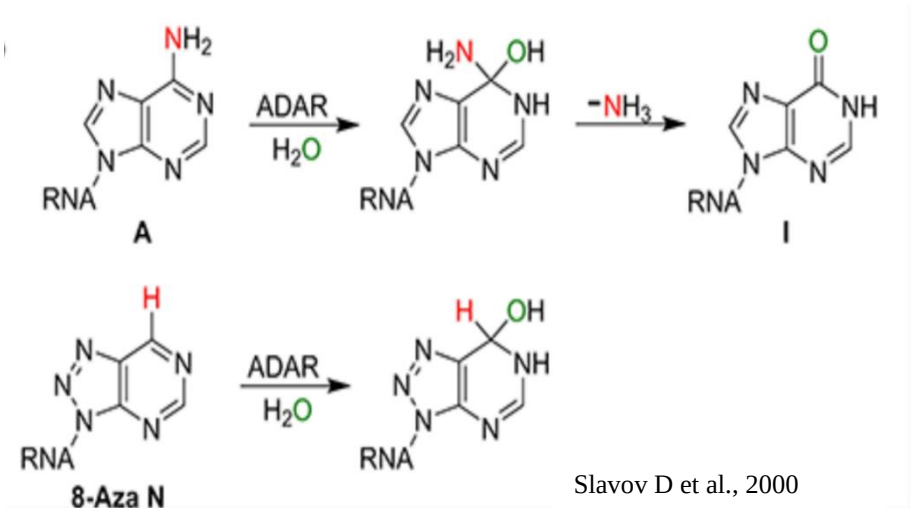
https://image.slidesharecdn.com/rnaediting-171031063231/95/rna-editing

Fig 2. Chemical structural differences of Adenosine (A), Inosine (I) and Guanosine (G)



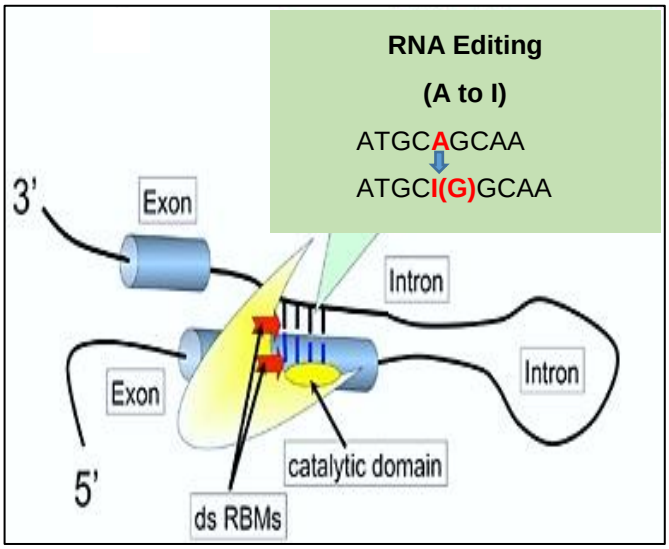
Slotkin et al., 2013

Fig 3. Different types of the ADAR deaminase



Slavov D et al., 2000

Fig 4. RNA editing by ADAR (Adenosine Deaminase Acting on RNA)



Hideyama et al., 2011

Fig 5. Mechanism of RNA editing by ADAR1 (A to I conversion)

1.7 Apolipoprotein B mRNA editing enzyme, catalytic polypeptide (APOBEC):

The APOBEC group of cytidine deaminases has extended and veered although throughout vertebrate evolution and contemporarily consists of AID and five APOBECs, numbered from 1 to 5 (abbreviated A1 to A5). The family most likely developed from AID (or AICDA), which is thought to have been evolved from the Tad (tRNA adenosine deaminase)/ADAT2 (adenosine deaminase, tRNA Specific 2) tRNA-altering compounds before the vertebrate radiation.

Advancement has formed a quite certain capacity for APOBEC in adaptive immunity. APOBEC is exceptionally communicated in enacted B cells and is fundamental for immunizer proclivity development and expansion (56). By DNA editing of the genomic immunoglobulin (Ig) loci APOBEC triggers particular downstream process: class-switch recombination (CSR), somatic hypermutation (SHM), and, in certain vertebrates, gene alteration. Intriguingly, Ig hypermutation by APOBEC can likewise bring about an erasure of the entire substantial chain locus, causing B cell passing in a

procedure named locus suicidal recombination (57-59).

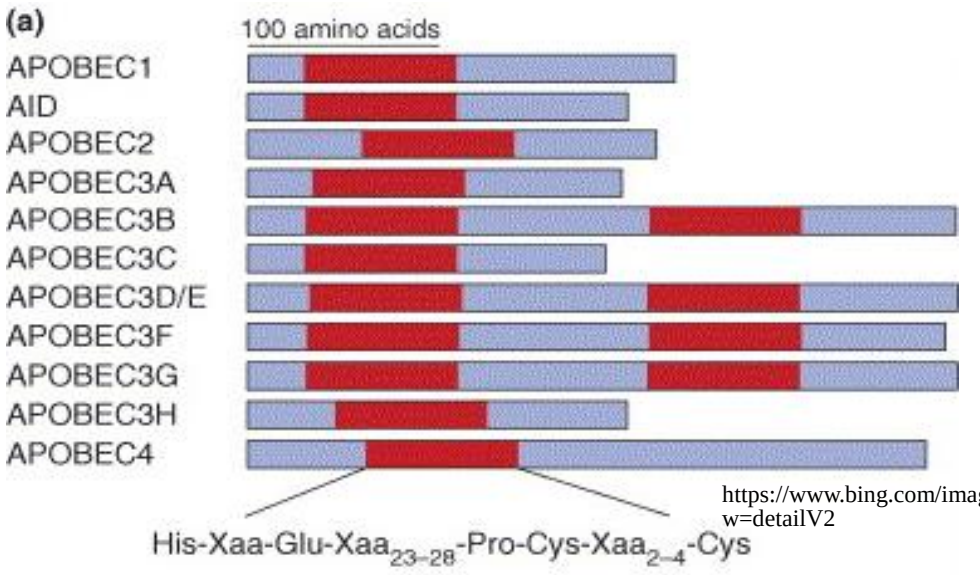
1.8 Types of the APOBEC and its function

APOBEC1, A1 is the first APOBEC to be found, works in humans as an RNA editor. A1 specifically C-to-U edits cytidine 6666 in ApoB mRNA, which encodes a key player in lipid transportation (60). This RNA editing, which happens explicitly in the small intestine in humans, makes an untimely stop codon that yields a short ApoB-48 peptide rather than the full-length 100 amino corrosive ApoB-100, the two of which are significant for lipid homeostasis (61). Thus, A1 operates similarly to ADAR enzymes by creating mRNAs with differential capacity from a single locus (62). Strikingly, A1 does not edit ApoB mRNA in amniotes, in spite of the fact that their A1 can alter genomic DNA when transfected to *E. coli* and confine retro-elements in *ex vivo* assays, proposing that these may reflect the genealogical function of A1. Like AID, this mutagenic capacity can be dangerous and overexpression of A1 can prompt to malignant growth.

APOBEC2, A2 is fundamentally expressed in heart and skeletal muscle in warm blooded animals (63, 64) and chickens (65). It is inessential for mouse development, survival, and fertility (66), but is needed for proper muscle growth (67, 68), for left–right specification in lower vertebrate embryogenesis (69), and for retina and optic nerve regeneration (70). The molecular capacity of A2 has been tricky in light of the fact that it lacks any identifiable cytidine deaminase activity. Ongoing work on its role in retina regeneration (71) has demonstrated that the zinc-coordinating domain but not deaminase activity are required to stimulate the binding of the Pou6f2 (POU class 6 homeobox 2) transcription factor to DNA. Regardless of whether cytidine deamination by A2 takes place in explicit contexts *in vivo*, perhaps working in combination with vital cofactors, is yet to be resolved.

APOBEC3, A3s comprising of seven paralogs in the human genome (A3A–D, A3F–H), are DNA editors that are popular for their roles in innate immunity as potent inhibitors of viral infections and retro-elements. However, they are they are associated with invulnerability in different manners: A3G contributes to cellular immunity by enhancing natural killer (NK) cell recognition. A recent and novel finding is that, upon enlistment by inflammation-associated factors, A3A edits the mRNAs of several qualities of genes, some associated with viral pathogenesis, in monocytes and macrophages (72). A3s can also directly edit nuclear or mitochondrial DNA and transfected plasmids. APOBEC3 (regardless of its proposed function in restricting endogenous retrotransposition) are inessential for mouse development, survival, or fertility.

The physiological functions of A4 and A5 are yet to know. A4 is expressed in testicles (73) and does not deaminate DNA in *E. coli* and yeast assays (74). A5 is present in non-mammalian tetrapods and its similitude to AID and A5 suggests that it can edit DNA.



https://www.bing.com/images/search?view=detailV2

Fig 6. Members of APOBEC family

Table 1. Human AID and APOBEC paralogs

Name	**Genomic location**	**Exons**	**Deamination domains**	**Expression**	**Cellular localization**	**Editing activity**	**Target**
AID	12q13	5	1	Activated B cells, testis	Mainly cytoplasmic, acts in the nucleus	DNA	Immunoglobulin gene
APOBEC 1	12q13.1	5	1	Small intestine	Cytoplasmic/ nuclear, acts in the nucleus	RNA, DNA	Apolipoprotein B m RNA
APOBEC 2	6p21	3	1	Skeletal muscle, heart	Cytoplasmic/ nuclear	Unknown	Unknown
APOBEC 3 A	22q13.1	5	1	Keratinocytes, blood	Cytoplasmic/ nuclear	DNA	Adeno-associated virus, retrotransposons
APOBEC 3B	22q13.1	8	2	Intestine, Uterus, mammary gland, keratinocytes, other	Predominantly nuclear	DNA	Retroviruses, retrotransposons, HBV
APOBEC 3C	22q13.1	4	1	Many tissues	Cytoplasmic/ nuclear	DNA	Retroviruses, retrotransposons, HBV
APOBEC 3DE	22q13.1	7	2	Thyroid, spleen, blood	Unknown	DNA	Retroviruses
APOBEC3 F	22q13.1	8	2	Many tissues	Cytoplasmic	DNA	Retroviruses, retrotransposons, HBV
APOBEC 3 G	22q13.1	8	2	Many tissues, T cells	Cytoplasmic	DNA	Retroviruses, retrotransposons, HBV
APOBEC 3H	22q13.1	5	1	Blood, thymus, thyroid, placenta	Unknown	DNA	Retroviruses
LOC196469*	12q23	1	2	Pseudogene	-	-	-
APOBEC4	Iq25.3	2	1	Testis	Unknown	Unknown	Unknown

1.9 Site Directed RNA editing methods:

Site directed RNA editing is a method to recode genetic information at the RNA level

(16). The approach is based on the enzymatic conversion of the adenosine (A) to inosine (I). In many biological or biochemical processes including translation, inosine are interpreted as guanosine (Figure 2), thus A-to-I RNA editing allows the recoding of amino acids, splice elements, miRNAs and miRNA binding sites among others (17). Recently a new method has been developed named the site directed RNA editing that employs engineered deaminases in combination of the short guide RNAs to recode a single adenosine bases at specific sites in any user defined transcripts (18, 19). Due to the usage of the guide RNAs the target selection and specificity is easily and rationally programmed based on the simple Watson and Crick base pairing rules (21, 22). A single base change in the genome or mRNA can cause disease. Some genetic diseases caused by point mutations are shown in Table 1. If I can restore mutated RNAs, then these diseases can be treated.

Table 2. T-to-C and G-to-A point mutated diseases

No.	Disease state	Gene Symbol	Base change	Amino acid	Codon	Author	Journal	Vol	Page	Year
1.	ADA deficiency	ADA	CTG-CCG	Leu pro	107	Hirschhorn	PNAS	87	6171	1990
2	ADA Deficiency	ADA	AAA-AGA	Lys-Arg	80	Valeria	EMBO J	5	113	1986
3	Adrenal Hyperplasia	CA21HB	AAC-AGC	Asn-Ser	494	Rodrigues	EMBO J	6	1653	1987
4	APRT Deficiency	ART	ATG-ACG	Met-Thr	136	Hidaka	JCI	81	945	1988
5	Albinism, Ocul (1)	TYR	GAC-GGC	Asp-Gly	42	King	MBM	8	19	1991
6	Albumin Komagome 2	ALB	CAT-CGT	His-Arg	128	Madison	PNAS	88	9853	1991
7	Aldolase A defic.	ALDA	GAT-GGT	Asp-Gly	128	Kishi	PNAS	84	8623	1988
8	Amyloid Polyneur	PALB	TAC-TGC	Tyr-Cys	114	Ueno	BBRC	169	143	1990
9	Amyloid Prealbumin	PALB	GTG-GCG	Val-Ala	30	Johns	CLIN GENET	41	70	82
10	Androgen insens. syn.	AR	TAC-TGC	Tyr-Cys	761	McPhaul	JCI	87	1413	1991
11	Antithrombin III def.	AT3	TTC-TCC	Phe-Ser	402	Olds	TH	65	670	1991
12	Antitrypsin ∝ 1 def.	PI	CTC-CCG	Leu-Pro	41	Takahashi	JBC	263	15528	1988
13	Antitrypsin ∝1 def.	PI	CTC-GCG	Val-Ala	213	Nukiwa	JBC	261	15989	1986
14	Chr. Granulomat. dis.	CYBB91	CAT-CGT	His-Arg	101	Bolscher	BlOOD	77	2482	1991
15	Cystic Fibrosis	CFR	TAT-TGT	Tyr-Cys	913	Vidaud	HUM	85	446	1990
16	Elliptocytosis	SPTA	CTC-CCG	Leu-Pro	207	Gallagher	JCI	89	892	1992
17	Elliptocytosis	SPTA	AAG-AGG	Lys-Arg	48	Floyd	BLOOD	78	1364	1991

18	Epidermolysis Bull	KRT14	CTG-CCG	Leu-Pro	384	Bonfias	Science	254	1202	1991
19	G6PD Deficiency	G6PD	CAC-CGC	His-Arg	32	Chao	NAR	19	6056	1991
20	G6PD Deficiency	G6PD	CTG-CCG	Leu-pro	968	beutler	BLOOD	74	2550	1989
21	Galactosaemia	GALT	CTG-CCG	Leu-Pro	195	Reichardt	GENOMICS	12	596	1992
22	Galactosaemia	GALT	CAG-CGG	Gln-Arg	188	Reichardt	AJHG	49	860	1991
23	Gangliosidosis GM1	GLB1	ATC-ACC	Ile-Thr	51	Yoshida	AJHG	49	435	1991
24	Gangliosidosis GM1	GLB1	TAT-TGT	Tyr-Cys	316	Yoshida	AJHG	49	435	1991
25	Gaucher's disease (1)	GBA	AAC-AGC	Asn-Ser	370	Tsuji	PNAS	85	2349	1988
26	Gaucher's disease (2)	GBA	CTG-CCG	Leu-Pro	444	Tsuji	PNAS	316	570	1987
27	Gyrate atrophy	OAT	CTT-CCT	Leu-Pro	402	Mitchell	PNAS	86	197	1989
28	HPRT deficiency	HPRT	CTA-CCA	Leu-Pro	40	Davidson	JCI	84	342	1989
29	HPRT deficiency	HPRT	ATT-ACT	Ile-Thr	41	Davidson	AJHG	48	951	1991
30	HPRT deficiency	HPRT	ATG-ACG	Met-Thr	56	Skopeckn	HUM GENET	85	111	1990
31	HPRT deficiency	HPRT	TTG-TCG	Leu-Ser	130	Gibbs	PNAS	86	1919	1989
32	HPRT deficiency	HPRT	ATT-ACT	Ile-Thr	131	Davidson	AJHG	48	951	1991
33	HPRT deficiency	HPRT	GAT-GGT	Asp-Gly	52	Lightfoot	HUM GENET	88	695	1992
34	HPRT deficiency	HPRT	ATT-ACT	Ile-Thr	182	Tarle	GENOMICS	10	499	1991
35	HPRT deficiency	HPRT	GAT-GGT	Asp-Gly	200	Davidson	JBC	264	20	1989
36	HPRT deficiency	HPRT	CAT-CGT	His-Arg	203	Tarle	GENOMICS	10	499	1991
37	Haemoglobin	HBB	CAT-CGT	His-Arg	117	Kutlar	HUM	86	591	1991
38	Haemolytic Anaemia	PGK	CTG-CCG	Leu-Pro	88	Maeda	BLOOD	77	1871	1991
39	Haemophilia A	F8	TTC-TCC	Phe-Ser	293	Higuchi	PNAS	88	7405	1991
40	Haemophilia A	F8	TTG-TCG	Leu-Ser	2166	Levinson	AJHG	46	53	1990
41	Haemophilia A	F8	GAA-CGA	Glu-Gly	272	Youssoufia	AJHG	42	867	1988
42	Haemophilia A	F8	AAA-AGA	Lys-Arg	425	Higuchi	PNAS	88	7405	1991

43	Haemophilia A	F8	TAT-TGT	Tyr-Cys	473	Higuchi	PNAS	88	7405	1991
44	Haemophilia A	F8	GAT-GGT	Asp-Gly	542	Higuchi	PNAS	88	7405	1991
45	Haemophilia A	F8	TAT-TGT	Tyr-Cys	1680	Traystman	GENOMICS	6	293	1990
46	Hepatic lipase def.	HL	AAT-AGT	Asn-Ser	193	Hegele	BBRC	179	78	1991
47	Insulin Resistance	INSR	CTG-CCG	Leu-Pro	233	Klinkham	EMBO J	8	2503	1989
48	Isovaleric Acidaemia	IVD	CTA-CCA	Leu-Pro	13	VOckley	AJHG	49	147	1991
49	LDLR deficiency	LDLR	TAT-TGT	Tyr-Cys	807	Davis	CELL	45	15	86
50	Laron dwarfism	GHR	TTT-TCT	Phe-Ser	96	Amselem	NEJM	321	989	1989
51	Leprechaunism	INSR	CAC-CGC	His-Arg	209	Kadowaki	JCI	86	254	1990
52	Leukocyte adhes. Def.	LFA1	CTA-CCA	Leu-Pro	149	Wardlaw	JEM	172	335	1990
53	Lipoprt. lipase def.	LPL	ATT-ACT	Ile-Thr	194	Henderson	JCI	87	2005	1991
54	Lipoprt. lipase def.	LPL	GAT-GGT	Asp-Gly	158	Ma	JBC	267	1918	1992
55	LCAM deficiency	LCAM	AAT-AGT	Asn-Ser	351	Nelson	JBC	267	3351	1992
56	MCAD deficiency	MCAD	ATA-ACA	Ile-thr	375	Yokota	AJHG	49	1280	1991
57	Marfan syndrome	COL1A2	CAG-CCG	Gln-Arg	618	Phillips	JCI	86	1723	1990
58	Methaemoglobin	DIA1	CTG-CCG	Leu-Pro	148	Katsube	AJHG	48	799	1991
59	Methylmalonicacid	MCM	CAT-CGT	His-Arg	532	Crane	JCI	89	385	1992
60	Neurofibromatosis (1)	NF1	CTC-CCG	Leu-Pro		Cawthon	CELL	62	193	1990
61	OTC deficiency	OTC	CTA-CCA	Leu-Pro	45	Grompe	AJHG	48	212	1991
62	OTC deficiency	OTC	CTT-CCT	Leu Pro	111	Grompe	AJHG	48	212	1991
63	Phenylketonuria	PAH	TTG-TCG	Leu-Ser	48	Konecki	HUM GENET	87	389	1991
64	Phenylketonuria	PAH	TTG-TCG	Leu-Ser	255	Hofman	AJHG	48	791	1991
65	Phenylketonuria	PAH	CTG-CCG	Leu-Pro	311	Licht-k	BIOCHEM	27	2881	1988
66	Phenylketonuria	PAH	TAT-TGT	Tyr-Cys	204	Wang	GENOMICS	10	449	1991

67	Phenylketonuria	PAH	GAA-GGA	Glu-Gly	221	Konecki	HUM GENET	87	389	1991
68	Phenylketonuria	PAH	TAC-TGC	Tyr-Cys	414	Okano	NEJM	324	1232	1991
69	Pompe disease	GAA	ATG-ACG	Met-Thr	318	Zhong	AJHG	49	635	1991
70	Retinitis Pigmentosa	RDS	CTG-CCG	Leu-Pro	185	Kajiwara	MATURE	354	480	1991
71	Retinitis Pigmentosa	RHO	TAC-TGC	Tyr-Cys	178	Sung	PNAS	88	6481	1991
72	Retinitis Pigmentosa	RHO	GAC-GGC	Asp-Gly	190	Sung	PNAS	88	6481	1991
73	Ster.18-hydrox. Def.	CYP18	GTG-GCG	Val-Ala	386	Mitsuuchi	BBRC	182	974	1992
74	Thalassaemia ∝	HBA2	ATG-ACG	Met-Thr	-1	Piratsu	JBC	259	12315	1984
75	Thalassaemia ∝	HBA2	CTG-CCG	Leu-Pro	125	Goossens	NATURE	296	854	1982
76	Thalassaemia ∝	HBB	CTG-CCG	Leu-Pro	110	Kobayshi	BLOOD	70	1688	1987
77	Thalassaemia ∝	HBD	CTG-CCG	Leu-Pro	141	Trifillis	BLOOD	78	3298	1991
78	Wilm` tumor	WT1	GAC-GGC	Asp-Gly	396	Pelletier	CELL	67	437	1991

1.10 Mechanism of A to I editing:

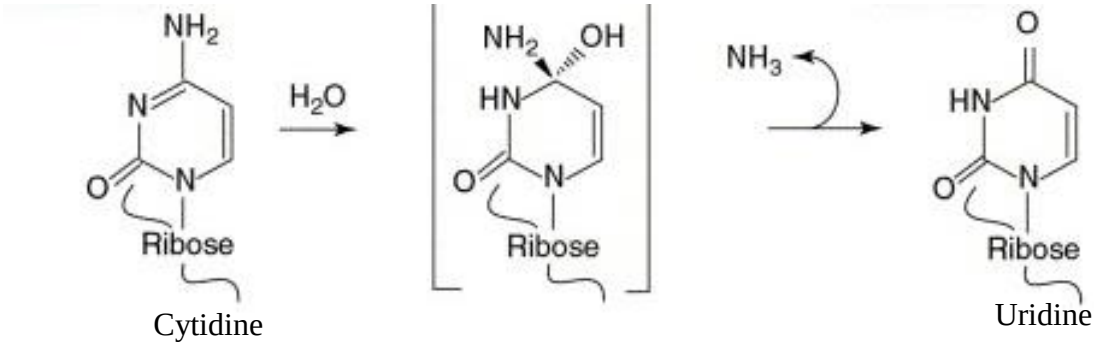
ADARs target the double stranded RNA which are formed both intra-molecularly and inter-molecularly. Editing occurs via the hydrolytic deamination at C6 of adenosine (A), changing it to inosine (I).

Cellular machineries read inosine (I) as guanosine (G) during the process of the translation, making the effect of A to I editing similar to an A to G substitution (Figure 5). This expands the transcriptome when editing occurs in coding mRNA sequences, in start and stop codons or at splice sites. However most known ADAR substrates are noncoding sequences, microRNAs (miRNAs) and the dsRNA formed from inverted Alu repeats. Binding and editing are generally not sequence specific (51), though ADAR 1, (52) and ADAR 2, (53) appear to have distinct 5 to 3 neighbor important in ADARs

recognition of its substrates (54, 55).

1.11 Mechanism of C-to-U editing:

C-to-U RNA editing is an important mechanism for amplifying mammalian genetic diversity in a regulated manner (Figure 7). Key to the success of this adaptation is the ability to define and limit access of the machinery to avoid enzymatic modifications within unintended targets. The identification of the core components of the apoB RNA editing holoenzyme and the ability to examine the role of new candidate genes that represent elements of the larger complex will likely reveal further functions in RNA metabolism. Establishing functional links between these distinct events should represent an exciting challenge for future years. C-to-U DNA editing creates U:G mispairs that, upon replication, lead to fixation of C-to-T or G-to-A transitions, depending on the strand edited. The U:G sites can also undergo base excision or mismatch repair, whose dysfunction can lead to both transitions or trans-versions at these sites.



https://www.bing.com/images/search?view=detailV2&ccid=8cyXXthy&id=5E8E78AD08CA670

Fig 7. Chemical modification during the conversion of Cytidine (C) to Uridine (U), which happens through a deamination process

Evidence has accumulated indicating that deamination in RNA is catalyzed by a superfamily of RNA-dependent deaminases. APOBEC1, the catalytic subunit of the complex that edits ApoB mRNA, was the first CDAR (Cytidine Deaminase Acting on RNA) cloned and, along with auxiliary proteins, forms a multisubunit complex that is

required for editing.

1.12 MS2 system

MS2 tagging is a technique which is based on the natural interaction between the MS2 bacteriophage coat protein with a stem-loop structure from the phage genome (75), which is used for biochemical purification of RNA-protein complexes and partnered to GFP expression for the detection of RNA in the living cells (76). More recently, the diversification of the system has been increased. The technique has been used to monitor the appearance of RNA in living cells, at the site of transcription, or simply by observing the changes in RNA number in the cytoplasm. There is another similar type bacteriophage derived tethering agent that is Lambda-N system. Both MS2 and Lambda-N system has the same small 12.2 kDa protein structure. Instead of its higher efficacy there are some disadvantages of the Lambda-N system. Such as: Binding affinity is 10^{-8}, whereas for MS2 system it is 10^{-9} when AUUA used or in case of AUCA even 10^{-10}. Off target effect in case of Lambda-N system is higher than the MS2 system and due to the higher molecular weight of the construct the delivery of the system is bit more difficult than MS2 system (86). For these reasons MS2 system is being more preferable for RNA editing (87, 88).

Start with single-stranded RNA, and create a pattern of stem-loop structures by adding copies of the MS2 RNA-binding sequences to a noncoding region. The MS2 protein must be fused with GFP and bonded to an mRNA, a complex that contains the RNA-binding sequence copies of the MS2. The MS2-GFP fusion protein is expressed by transferring it to a cell with a plasmid 5 (Robert Singer's lab). The signal encodes within RNA and the signal presences of the nuclear localization signal (NLS) within GFP-MS2, there are two signals those are introduced from the EGFP-MS2-RNA

complexes.

MS2 biotin-tagged RNA affinity purification (MS2-BioTRAP) is one of the *in vivo* methods, which is specified for the identification of the protein-RNA interactions (77). Both the RNA that tagged with MS2 and the MS2 protein tag are expressed, and then, the affinity is used to help the process of identifying protein-RNA interactions. The MS2 bacteriophage has the ability to bind its coat protein with the RNA stem loop. This tight association helps to monitor the appearance of the Nucleic acids within the living cells (Figure 8) and (Figure 10).

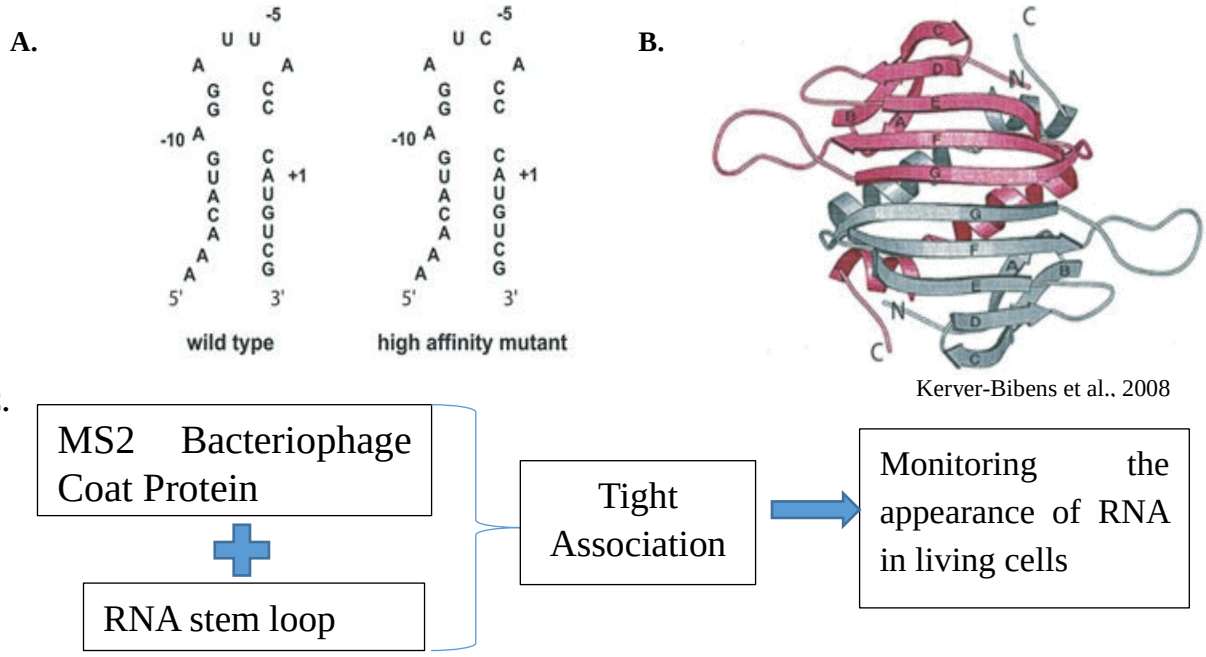
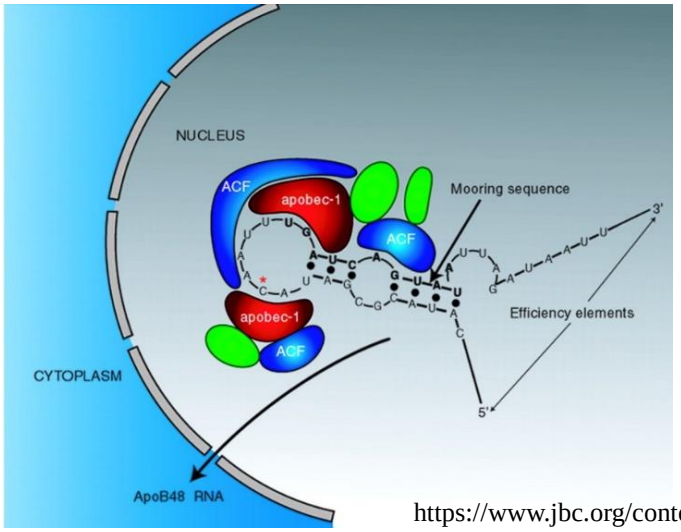


Fig 8. A. MS2-RNA stem-loop wild-type, **B.** change in few nucleotide result higher binding mutant, The numbering is from the first base of the AUG which is the initiation codon of the replicase mRNA C. Antiparallel oriented MS2 protein dimer with 10 β-sheets., The association between two MS2 coat protein monomers (red and grey). The β-sheets are named A to G, and the N- and C-termini of each monomer is indicated.
C. Schematic representation of the MS2 system

1.13 Mechanism of APOBEC for RNA editing:

ApoB RNA editing changes a CAA to a UAA stop codon, generating a truncated protein, apoB48 (79). ApoB RNA editing has important effects on lipoprotein metabolism, and its emergence defines distinct pathways for intestinal and hepatic lipid transport in mammals (79). C-to-U editing of apoB RNA requires a single-strand template with well-defined characteristics in the immediate vicinity of the edited base, as well as protein cofactors that assemble into a functional complex referred to as a holoenzyme or editosome. This functional complex includes a minimal core composed of APOBEC1, the catalytic deaminase, and a competence factor, APOBEC1 complementation factor (ACF), that functions as an adaptor protein by binding both the deaminase and the RNA substrate (Figure 9). The interaction of these protein components and their higher order interactions with the nuclear transcript illustrates the complexity of site-selectivity in C-to-U RNA editing (78).



https://www.jbc.org/content/278/3/1395/F1.large.jpg

Fig 9. C-to-U RNA editing of apolipoprotein B. The model for a 35-nucleotide region of apoB RNA flanking the edited base (asterisk) is shown. A schematic representation illustrates APOBEC1 (red) and ACF (blue) binding to RNA both 5' and 3' of the edited base and depicts the presence of additional proteins that may modulate assembly of the holoenzyme (green). Note that the stoichiometry of APOBEC-1 and ACF molecules with respect to the active enzyme is unknown. The model emphasizes the role of both cis-acting elements within the vicinity of the edited base (mooring sequence is bolded) and the requirement for an optimal structure, conferred by both 5' and 3' efficiency elements.

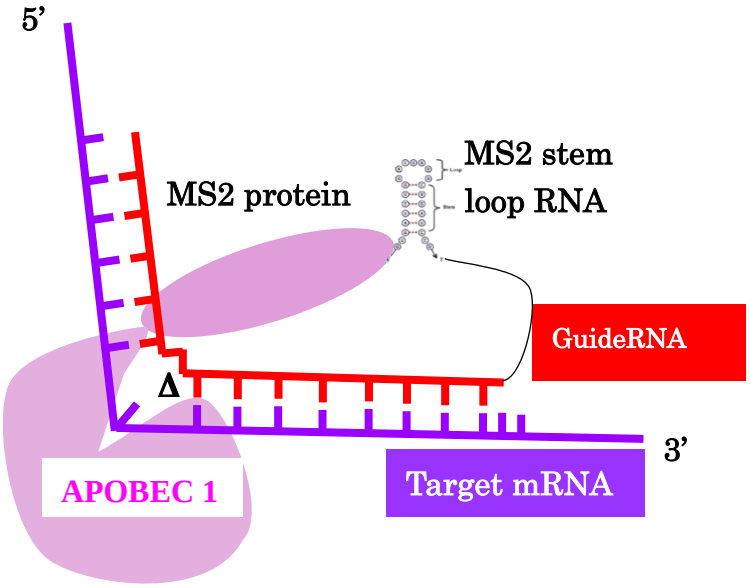


Fig 10. Mechanism for the RNA editing using APOBEC1 and guideRNA, having a mismatch at the target position in the guideRNA

1.14 Aim of the study:

In case of the gene therapy whole gene or transcript is targeted but my aim to correct point mutations in case of G-to-A mutation that result stop codon and in case of T-to-C mutation BFP. Ultimately those defective codons produce defective or non-functional proteins and that cause diseases. The defective codons can be altered or corrected to the desired one or another read through codon by base substitution. The efficacy of the restoration and also the effective application of the developed artificial system as a therapy is very important. Furthermore compared to other editing tools the efficacy of the MS2 system for RNA editing is also needed to establish especially for C-to-U RNA editing, as this is the first time application of MS2-APOBEC approach. By considering all the above mentioned points the major aims of this study are-

i) Development of an artificial deaminase complex using ADAR1 and APOBEC1 along with MS2 system, to restore the G-to-A (TAA-ochre stop codon) and T-to-C (CCC-BFP) mutations

ii) Finding out the most effective promoter combination for increasing the efficacy of the restoration in case of C-to-U editing

iii) Application of the developed artificial deaminase system *in vivo* on real model mouse for checking the efficacy of the system in respect to gene therapy

1.15 Expected impact of the study:

Development of the artificial deamination system for the treatment of the genetic diseases caused by the G-to-A or T-to-C point mutations and making the treatment of genetic diseases available and under the capacity of mass people.

1.16 REFERENCES

1. Heidenreich M. and Zhang, F. Applications of CRISPR-Cas systems in neuroscience. Nat Rev Neurosci, 17: 36–44, (2016).
2. Kim Y.G., Cha J. and Chandrasegaran S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proc Natl Acad Sci, 93: 1156–1160, (1996).
3. Porteus M.H. and Carroll D. Gene targeting using zinc finger nucleases. Nat Biotechnol, 23: 967–973, (2005).
4. Zhang F., Le C., Simona L., Sriram K., George M.C., and Paola A. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. Nat Biotechnol, 29: 149–153, (2011).
5. Mali P., Luhan Y., Kevin M.E., Joh A., Marc G., James E.D., Julie, E.N. and George M. C. RNA-guided human genome engineering via Cas9. Science, 339: 823–826, (2013).
6. Araki M. and Ishii T. Providing Appropriate Risk Information on Genome Editing for Patients. Trends Biotechnol, 34: 86–90, (2016).
7. 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).
8. Gott J.M., and Emeson, R.B. Annu. Rev. Genet. 34: 499–531, (2000)
9. .Gerber A.P., and Keller W. Trends Biochem. Sci. 26: 376–384, (2001)
10. Zhang F. et al. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. Nat Biotechnol, 29: 149–153, (2011).
11. Mali P. et al. RNA-guided human genome engineering via Cas9. Science, 339: 823–826, (2013).
12. Vogel P. and Stafforst, T. Site-directed RNA editing with antagomir deaminases—A

tool to study protein and RNA function. ChemMedChem, 9: 2021–2025, (2014).

13. Nishikura K. Functions and regulation of RNA editing by ADAR deaminases. Annu. Rev. Biochem., 79: 321–349, (2010).
14. Stafforst T. and Schneider M.F. An RNA Deaminase conjugates electively repairs point mutations. Angew. Chem. Int. Ed. 51: 11166–11169, (2012).
15. Montiel-Gonzalez M.F., Guillermo I., Yudowski A., and Rosenthal J.J.C. Correction of mutations within the cysticfibrosistransmembraneconductanceregulatorbysite-directedRNAediting. Proc. Natl. Acad. Sci. USA. 110: 18285–18290. (2013).
16. 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: 6267–6271, (2014).
17. 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: 15875–15881, (2015).
18. Bass B.L. RNA editing by adenosine deaminases that act on RNA. Annu. Rev. Biochem., 71: 817–846, (2002).
19. Bass B.L., Nishikura K., Keller W., Seeburg P.H., Emeson R.B., O'Connell M.A., Samuel C.E. and Herbert A. A standardized nomenclature for adenosine deaminases that act on RNA. RNA, 3: 947–949, (1997).
20. Nishikura K. Editor meets silencer: crosstalk between RNA editing and RNA interference. Nat. Rev. Mol. Cell Biol., 7: 919–931, (2006).
21. Burns C.M., Chu H., Rueter S.M., Hutchinson L.K., Canton H., Sanders-Bush E., and Emeson R.B. Regulationofserotonin-2C receptor G-protein coupling by RNA

editing. Nature, 387: 303–308. (1997).

22. Higuchi M., Single F.N., Kohler M., Sommer B., Sprengel R. and Seeburg P.H. RNA editing of AMPA receptor subunit GluR-B: a base-paired intron-exon structure determines position and efficiency. Cell, 75: 1361–1370, (1993).
23. Tonkin L. A., and Bass B. L. Mutations in RNAi rescue aberrant chemotaxis of ADAR mutants. Science, 302: 1725, (2003).
24. Rueter S.M., Dawson T.R., and Emeson R.B. Regulation of alternative splicing by RNA editing. Nature, 399: 75–80, (1999).
25. Scadden A.D.J., and Smith C.W.J. RNAi is antagonized by A→I hyper-editing. EMBOJ, 2: 1107-1111, (2001).
26. Yang W., Chendrimada T.P., Wang Q., Higuchi M., Seeburg P.H., Shiekhattar R., and Nishikura K. Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. Nat. Struct. Mol. Biol., 13: 13–21, (2006).
27. Athanasiadis, A., Rich, A. & Maas, S. Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. PLoS Biol. 2: 391, (2004).
28. Jin Y., Zhang W. and Li Q. Origins and evolution of ADAR-mediated RNA editing. IUBMB Life, 61: 572-578, (2009).
29. Kawahara Y., Ito K., Sun H., Aizawa H., Kanazawa I. and Kwak S. Glutamate receptors: RNA editing and death of motor neurons. Nature, 427: 801-801, (2004).
30. Li M., Yang L., Li C., Jin C., Lai M., Zhang G., Hu Y., Ji J., and Yao Z. Mutational spectrum of the ADAR1 gene in dyschromatosis symmetrica hereditaria. Arch. Dermatol. Res., 302: 469-476, (2010).
31. Morabito M.V., Abbas A.I., Hood J.L., Kesterson R.A., Jacobs M.M., Kump D.S., Hachey D.L., Roth B.L. and Emeson R. B. Mice with altered serotonin 2C receptor

RNA editing display characteristics of Prader–Willi syndrome. Neurobiol. Dis., 39: 169–180, (2010).

32. Silberberg G., Lundin D., Navon R. and Őhman M. Deregulation of the A-to-I RNA editing mechanism in psychiatric disorders. Hum. Mol. Gen., 21: 311–321, (2012).
33. Cenci C., Barzotti R., Galeano F., Corbelli S., Rota R., Massimi L., Di Rocco C., O'Connell M.A. and Gallo A. Down-regulation of RNA editing in pediatric astrocytomas: ADAR2 editing activity inhibits cell migration and proliferation. J. Biol. Chem., 283: 7251–7260, (2008).
34. Paz N., Levanon E.Y., Amariglio N., Heimberger A.B., Ram Z., Constantini S., Barbash Z.S., Adamsky K., Safran M. and Hirschberg A. Altered adenosine-to-inosine RNA editing in human cancer. Genome Res., 17: 1586–1595, (2007).
35. Zhang X.J., He P.P., Li M., He C.D., Yan K.L., Cui Y., Yang S., Zhang K.Y., Gao M. and Chen J.J. Seven novel mutations of the ADAR gene in Chinese families and sporadic patients with dyschromatosis symmetrica hereditaria (DSH). Hum. Mutat. 23: 629–630, (2004).
36. Rice G.I., Kasher P.R., Forte G.M.A., Mannion N.M., Greenwood S.M., Szynkiewicz M., Dickerson J.E., Bhaskar S.S., Zampini M. and Briggs T.A. Mutations in ADAR1 cause Aicardi-Goutieres syndrome associated with a type I interferon signature. Nat. Genet., 44: 1243–1248, (2012).
37. 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: 15875–15881, (2015).
38. Barraud P., and Allain F.H. ADAR proteins: double-stranded RNA and Z-DNA

binding domains. Curr. Top. Microbiol. Immunol., 353: 35–60, (2012).

39. Ryter J.M., and Schultz S.C. Molecular basis of double-stranded RNA-protein interactions: structure of a dsRNA-binding domain complexed with dsRNA. EMBO J., 17: 7505–7513, (1998).
40. Vogel P., and Stafforst T. Site-directed RNA editing with antagomir deaminases-A tool to study protein and RNA function. ChemMedChem, 9: 2021–2025, (2014).
41. Slotkin W., and Nishikura, K. Adenosine-to-inosine RNA editing and human disease. Genome Med. 5: 105, (2013).
42. Stefl R., Xu M., Skrisovska L., Emeson R.B. and Allain, F.H. Structure and specific RNA binding of ADAR2 double-stranded RNA binding motifs. Structure, 14: 345–355, (2006).
43. Stephens O.M., Haudenschild B.L. and Beal P.A. The binding selectivity of ADAR2's dsRBMs contributes to RNA-editing selectivity. Chem. Biol., 11: 1239–1250, (2004).
44. Bass B.L. RNA editing and hypermutation by adenosine deamination. Trends Biochem. Sci., 22: 157–162, (1997)
45. Polson A.G. and Bass B.L. Preferential selection of adenosines for modification by double-stranded RNA adenosine deaminase. EMBO J., 13: 5701–5711, (1994).
46. Lehmann K.A. and Bass B.L. Double-stranded RNA adenosine deaminases ADAR1 and ADAR2 have overlapping specificities. Biochemistry, 39: 12875–12884, (2000).
47. Stefl R., Oberstrass F.C., Hood J.L., Jourdan M., Zimmermann M., Skrisovska L., Maris C., Peng L., Hofr C. and Emeson R.B. The solution structure of the ADAR2 dsrbm-RNA complex reveals a sequence-specific readout of the minor groove. Cell. 143: 225–237, (2010).

48. Muramatsu M., Kinoshita K., Fagarasan S., Yamada S., Shinkai Y., and Honjo T. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell., 102: 553-563 (2000).
49. Hwang J.K., Alt F.W., and Yeap L-S. Related mechanisms of antibody somatic hypermutation and class switch recombination. Microbiol. Spectr., 3(1): 1-35, (2015).
50. Chen Z., Wang J.H. Generation and repair of AID-initiated DNA lesions in B lymphocytes. Front. Med., 8: 201-216, (2014).
51. Peron S., Laffleur B., Denis-Lagache N., Cook-Moreau J., Tinguely A., Delpy L., Denizot Y., Pinaud E., and Cogne M. AID-driven deletion causes immunoglobulin heavy chain locus suicide recombination in B cells. Science, 336(6083):931-4, (2012).
52. Knisbacher B.A., Gerber D., and Levanon E.Y. DNA Editing by APOBECs: A Genomic Preserver and Transformer. Trends in genetics, 32(1): 16-28, (2016).
53. Daniels T.F., Killinger K.M., Michal J.J., Wright Jr. R.W., and Jiang Z. Lipoproteins, cholesterol homeostasis and cardiac health. Int. J. Biol. Sci., 5(5): 474-488, (2009).
54. Knisbacher B.A., and Levanon E.Y. DNA and RNA editing of retrotransposons accelerate mammalian genome evolution. Ann. N. Y. Acad Sci., 1341: 115-125, (2015).
55. Liao W., Hong S.H., Chan B.H., Rudolph F.B., Clark S.C., and Chan L. APOBEC-2, a cardiac and skeletal muscle-specific member of the cytidine deaminase supergene family. Biochem. Biophys. Res. Commun., 260: 398-404, (1999).
56. Anant S., Mukhopadhyay D., Sankaranand V., Kennedy S., Henderson J. O., and Davidson N.O. ARCD-1, an apobec-1-related cytidine deaminase, exerts a dominant

negative effect on C to U RNA editing Am. J. Physiol. Cell Physiol., 281: C1904-C1916, (2001)

57. Li J., Zhao X-L., Gilbert E.R., Li D-Y., Liu Y.-P., Yan W., Zhu Q., Wang Y-G., Chen Y., Tian K. APOBEC2 mRNA and protein is predominantly expressed in skeletal and cardiac muscles of chickens. Gene, 539: 263-269, (2014).
58. Mikl M.C., Watt I.N., Lu M. Reik W., Davies S.L., Neuberger M.S., and Rada C. Mice deficient in APOBEC2 and APOBEC3. Mol. Cell. Biol., 25: 7270-7277, (2005).
59. Sato Y., Probst H.C., Tatsumi R., Ikeuchi Y., Neuberger M.S., and Rada C. Deficiency in APOBEC2 leads to a shift in muscle fiber type, diminished body mass, and myopathy. J. Biol. Chem., 285: 7111-7118, (2010).
60. Etard C., Roostalu U., and Strahle U. Lack of Apobec2-related proteins causes a dystrophic muscle phenotype in zebrafish embryos. J. Cell Biol., 189: 527-539, (2010)
61. Vonica A., Rosa A., Arduini B., and Brivanlou A.H. APOBEC2, a selective inhibitor of TGFβ signaling, regulates left-right axis specification during early embryogenesis. Dev. Biol., 350: 13-23, (2011).
62. Powell C., Elsaeidi F., and Goldman D. Injury-dependent Müller glia and ganglion cell reprogramming during tissue regeneration requires Apobec2a and Apobec2b. J. Neurosci., 32:1096-1109, (2012).
63. Powell C., Cornblath E., and Goldman D. Zinc-binding domain-dependent, deaminase-independent actions of apolipoprotein B mRNA-editing enzyme, catalytic polypeptide 2 (Apobec2), mediate its effect on zebrafish retina regeneration. J. Biol. Chem., 289 (42): 28924–28941, (2014)

64. Sharma S., Patnaik S.K., Taggart R.T., Kannisto E.D., Enriquez S.M., Gollnick P., and Baysal B.E. APOBEC3A cytidine deaminase induces RNA editing in monocytes and macrophages. Nat. Commun., 6: 6881, (2015).
65. Rogozin I.B., Basu M.K., Jordan I.K., Pavlov Y.I. and Koonin E.V.. APOBEC4, a new member of the AID/APOBEC family of polynucleotide (deoxy) cytidine deaminases predicted by computational analysis. Cell Cycle, 4:1281-1285, (2005).
66. Lada A.G., Krick C.F., Kozmin S.G., Mayorov V.I., Karpova T.S., Rogozin I.B., and Pavlov Y.I. Mutator effects and mutation signatures of editing deaminases produced in bacteria and yeast. Biochemistry (Mosc.), 76: 131-146, (2011).
67. Johansson H.E., Liljas L., and Uhlenbeck O.C. RNA recognition by the MS2 phage coat protein. Sem Virol., 8 (3): 176–185, (1997).
68. Bertrand E., Chartrand P., Schaefer M., Shenoy S.M., Singer R.H. and Long R.M. Localization of ASH1 mRNA particles in living yeast. Molecular Cell, 2 (4): 437–45, (1998).
69. Hsu P.D., Lander E.S., and Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. Cell., 157 (6): 1262–78, (2014).
70. Blanc V., and Davidson O.N. C-to-U RNA Editing: Mechanisms Leading to Genetic Diversity. The Journal of Biological Chemistr. 278(3): 1395–1398, (2003).
71. Anant S., Davidson N.O. Curr. Opin. Lipidol. 12:159–165(2001)
72. Muramatsu M., Sankaranand V.S., Anant S., Sugai M., Kinoshita K., Davidson N.O., Honjo T. Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. J Biol Chem. 274: 18470-18476, (1999).

73. Navaratnam N., Morrison J.R., Bhattacharya S., Patel D., Funahashi T., Giannoni F., Teng B.B., Davidson N.O., Scott J. The p27 catalytic subunit of the apolipoprotein B mRNA editing enzyme is a cytidine deaminase. J Biol Chem., 268: 20709-20712, (1993).

74. Teng B., Burant C.F., Davidson N.O. Molecular cloning of an apolipoprotein B messenger RNA editing protein. Science., 260: 1816-1819, (1993). 10.1126/science.8511591.

75. Muramatsu M., Kinoshita K., Fagarasan S., Yamada S., Shinkai Y., Honjo T. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell., 102: 553-563, (2000).

76. Liao W., Hong S.H., Chan B.H., Rudolph F.B., Clark S.C., Chan L. APOBEC-2, a cardiac- and skeletal muscle-specific member of the cytidine deaminase supergene family. Biochem Biophys Res Commun., 260: 398-404, (1999).

77. Anant S., Henderson JO., Mukhopadhyay D., Navaratnam N., Kennedy S., Min J., Davidson N.O. Novel role for RNA-binding protein CUGBP2 in mammalian RNA editing. CUGBP2 modulates C to U editing of apolipoprotein B mRNA by interacting with apobec-1 and ACF, the apobec-1 complementation factor. J Biol Chem., 276: 47338-47351, (2001).

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

79. Madsen P., Anant S., Rasmussen H.H., Gromov P., Vorum H., Dumanski J.P., Tommerup N, Collins JE, Wright CL, Dunham I, MacGinnitie AJ, Davidson NO, and Celis JE. Psoriasis upregulated phorbolin-1 shares structural but not functional

similarity to the mRNA-editing protein apobec-1. J Invest Dermatol., 113: 162-169, (1999).

80. Wedekind J.E., Dance G.S., Sowden M.P., and Smith H.C. Messenger RNA editing in mammals: new members of the APOBEC family seeking roles in the family business. Trends Genet., 19: 207-216, (2003).
81. Conticello S.G., Thomas C.J., Petersen-Mahrt S.K., Neuberger M.S. Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. Mol Biol Evol., 22: 367-377, (2005).
82. Sheehy A.M., Gaddis N.C., Choi J.D., Malim M.H. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature., 418: 646-650, (2002).
83. Rogozin I.B., Basu M.K., Jordan I.K., Pavlov Y.I., Koonin E.V. APOBEC4, a new member of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases predicted by computational analysis. Cell Cycle., 4: 1281-1285, (2005).
84. George C.X., Gan Z., Liu Y., Samuel C.E. Adenosine Deaminases Acting on RNA, RNA Editing, and Interferon Action. J Interferon Cytokine Res. 2011; 31(1): 99–117
85. Mao S., Liu Y., Huang S., Huang X, Chi T. Site-directed RNA editing (SDRE): Off-target effects and their countermeasures. Journal of Genetics and Genomics. 2019; 46(11): 531-35.
86. Montiel-Gonzalez M.F., Quiroz J.F.D., Rosenthal J.J.C. Current strategies for Site-Directed RNA Editing using ADARs. Science Direct Methods; 156: 16-24, (2019).

87. Bhakta S., Azad T.A., Tsukahara T. Genetic code restoration by artificial RNA editing of Ochre stop codon with ADAR1 deaminase. Protein Engineering, Design and Selection, 31(12): 471-78, (2018).
88. Bhakta S., Tsukahara T. Artificial RNA editing with ADAR for genetic code restoration. Current Gene therapy, 20(1): 44-54, (2020)

CHAPTER II

A-to-I RNA editing by using ADAR1 artificial deaminase system for restoration of genetic code in Ochre (UAA) stop codon

2.1 INTRODUCTION

The techniques that regulate the functions and expressions of intracellular target genes, called genetic engineering. Those are used widely in basic researches, as well as medicinal and therapeutic applications (1, 6). In recent times, several genome editing technologies have made the manipulation of target genomic information attainable (6, 11, 15, 24, 12). In addition to programmable regulation, genome alteration can alter targeted cellular genes, thereby regulating the expression of target proteins (1). However, genome editing has some drawbacks, as modifications at other than target sites may be permanent and may affect genome sequences. By contrast, mismatches resulting from RNA editing will not be permanent, as they do not affect genome sequences. RNA editing may be preferable to genome editing for treatment purposes. I therefore utilized RNA editing to develop an artificial deaminase technique to restore the genetic code in stop codons (ochre). In our laboratory, we previously applied this RNA editing technique to amber (2) and opal stop codons. This study therefore applied this method to ochre stop codon to restore the genetic code.

Site directed RNA editing is a method to recode genetic information at the RNA level (22). The approach is based on the enzymatic conversion of the adenosine (A) to inosine (I). During many biological or biochemical processes, including translation, inosine (I) is interpreted as guanosine (G); thus A to I RNA editing replaces A with G residues, allowing the restoration of codes of amino acids, splice elements, miRNAs, and miRNA-binding sites (14). Site directed RNA editing is a new and recently developed method using engineered deaminases combined with short guide RNAs to recode single A residues at specific sites in any user defined transcript (2, 19, 25, 13). The use of guide RNAs results in the easy and rational programming of target selection and

specificity based on simple Watson and Crick base pairing rules (21, 5). MS2 tagging, which is based on natural interactions between the MS2 bacteriophage coat protein and a stem-loop structure from the phage genome (9), is used for biochemical purification of RNA-protein complexes and is coupled with green fluorescent protein (GFP) expression to detect RNA in living cells (3). The MS2 protein must be fused with GFP and bound to a mRNA, a complex that contains copies of the RNA-binding sequence copies of MS2 (Figure 11). With the help of this phenomenon of the MS2, the ADAR1 would come across the Guide RNA as coat protein is attached with the ADAR1 construct and Stem loop with the Guide RNA construct. MS2 biotin-tagged RNA affinity purification (MS2-BioTRAP) is one of the in vivo methods, which is specified for the identification of the protein-RNA interactions (7). RNA, in our laboratory, previously we restored an amber stop codon TAG to the readable codon TGG in enhanced green fluorescent protein (EGFP), which required only a single mutation (2, 25). The MS2 system is more frequently used than the Lambda N/B box systems to tether the protein to RNAs. To my knowledge, this was the first study showing that the MS2 system could be utilized in designing an RNA editing enzyme complex capable of targeting specific point mutations. Successful targeting of the amber stop codon was confirmed by JuLi fluorescence microscopy, which showed the expression of fluorescence, and polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP), which showed that TAG was restored. Sequencing and western blotting results also showed that the MS2 system successfully restored the genetic sequence of this TAG stop codon. But in case of Ochre stop codon I only presented the JuLi microscopic observation of the fluorescence in that paper (2). The MS2 system and a genetically encoded simple guide RNA to direct the deaminase domain of the RNA editing enzyme adenosine deaminase

acting on RNA (ADAR1) to specific targets may also effectively restore the genetic code of an ochre stop codon (TAA) (2). This study analyzed the effectiveness of this system in converting TAA to TGG, using confocal microscopy, PCR-RFLP analysis, and direct sequencing.

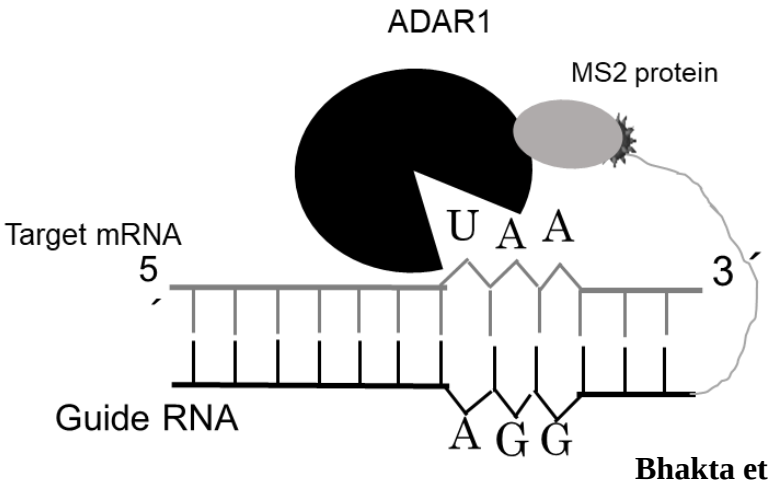


Fig 11. Schematic feature of the engineered ADAR1 (Black), MS2 protein (Grey), MS2 RNA (Dark ash) connected with Guide RNA. In the target RNA where the target Adenosine will be converted to Inosine (I) which is recognized as the Guanosine at the time of translation.

2.2 MATERIALS AND METHODS

2.2.1 Preparation of the plasmid constructs:

The ochre stop codon (TAA)-containing plasmid was constructed by site directed mutagenesis of the plasmid pcDNA-EGFP, with a TGG sequence at codon 189. TGG was converted to TAA (using a Quick change II site Directed Mutagenesis Kit (Agilent Technologies) and the specific primers 5'-AAGCTGCCCGTGCCCTAGCCCACCCTCGT-3' (forward) and 5'-ACGAGGGTGGGCTAGGGCACGGGCAGCTT-3' (reverse).

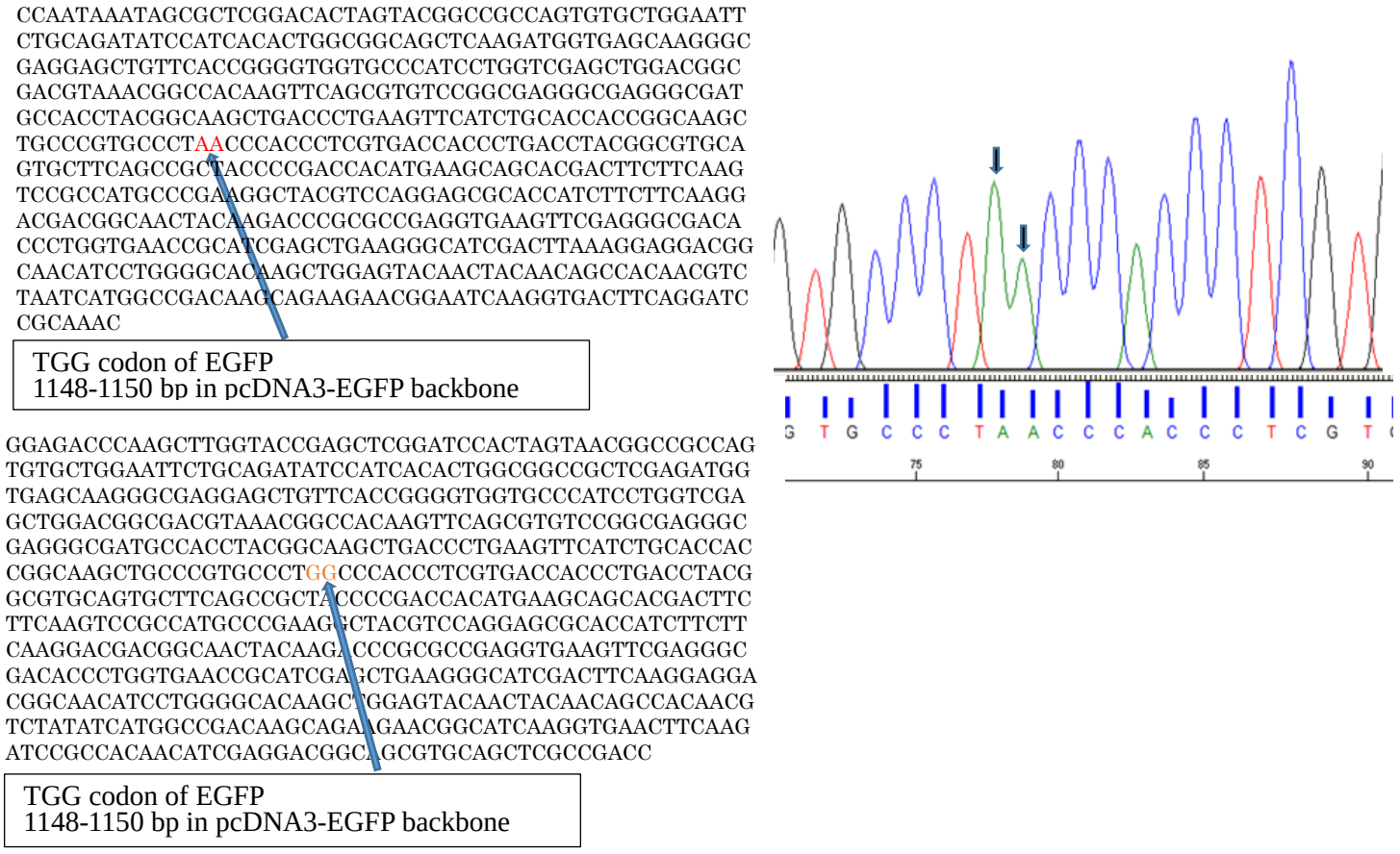


Fig 12. Sequence result of the target mRNA, EGFP containing Ochre stop codon (TAA)

2.2.2 Construction of ADAR1:

cDNAs encoding human ADAR1 and MS2-HB protein were inserted into the plasmid pCS2+MT following digestion with the restriction enzymes NcoI, XhoI, and XbaI (2). Plasmid DNA was extracted using a QIAGEN Minikit (QIAGEN) and cloned into Escherichia coli DH5∝, with its identity confirmed by sequencing.

2.2.3 Preparing guide RNA for directing ADAR1 to target:

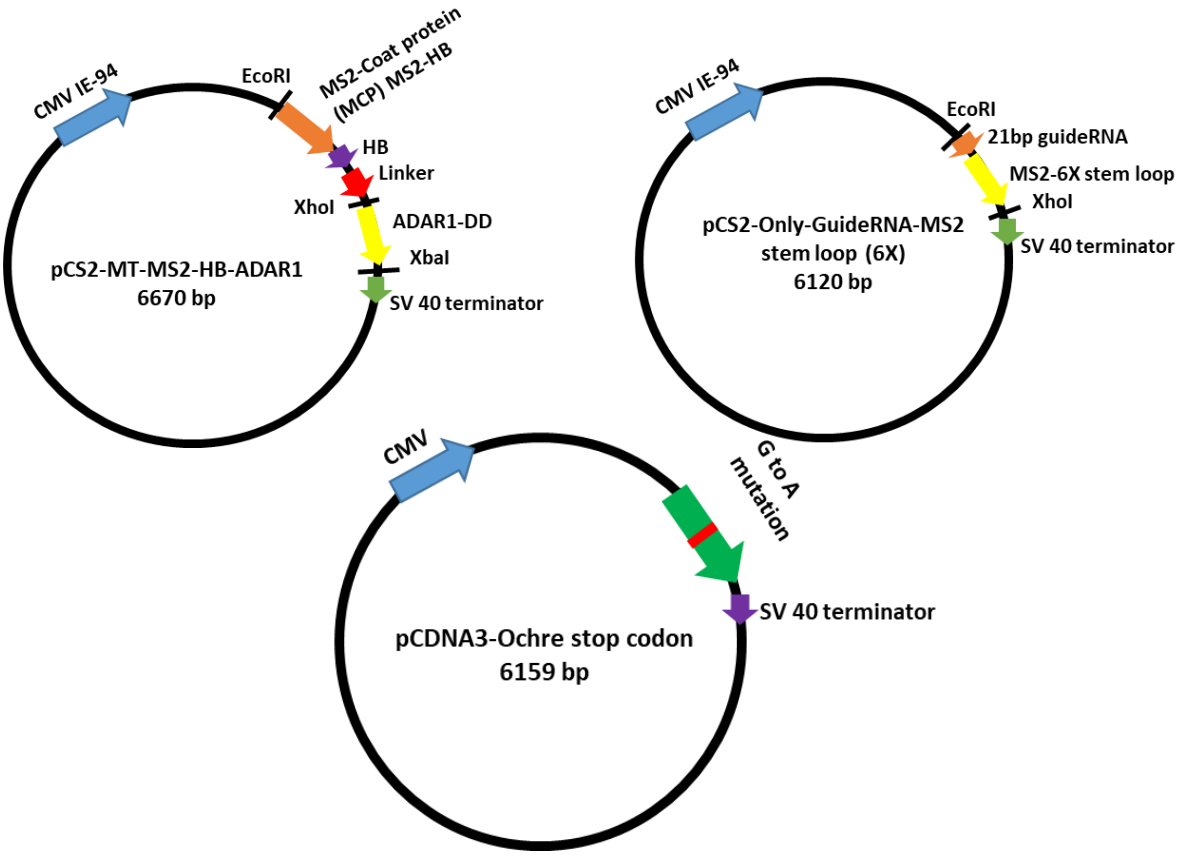
The 19-nucleotides containing guide RNA, which was complementary to the target mRNA site, was inserted either upstream

(5'-atcaGAATTCCACGGGAGGGGTGGGAGGGAATGGCCATGGGACGTCGAC-3')

or downstream

(5'attcCTCGAGGGAGGGTGGGGAGGGCACCGCAAATTTAAAGCGCTGAT-3') of MS2-RNA using either the forward or reverse primer (pCS2+Guide-MS2-RNA-Guide). The nucleotides atca and attc were inserted to identify these sequences by restriction enzyme digestion, the bold sequences represent the restriction enzyme sites, the sequences in italics represent the 19 bp guide, and the underlined sequences are the forward and reverse primers for MS2. Plasmid DNA was extracted using a QIAGEN Minikit (QIAGEN) and cloned into E. coli DH5∝, with its identity confirmed by

sequencing (Figure 13).



GAATTCCACGGGAGGGGTGGGAGGGAATGGCCATGGGACGTCGACGGACGTCGACCT
GAGGTAATTATAACCCGGGCCCTATATATGGATCCTAAGGTACCTAATTGCCTAGAAAACAT
GAGGATCACCCATGTCTGCAGGTCGACTCTAGAAAACATGAGGATCACCCATGTCTGCAG
TATTCCCGGGTTCATTAGATCCTAAGGTACCTAATTGCCTAGAAAACATGAGGATCACCCA
TGTCTGCAGGTCGACTCTAGAAAACATGAGGATCACCCATGTCTGCAGTATTCCCGGGTT
CATTAGATCCTAAGGTACCTAATTGCCTAGAAAACATGAGGATCACCCATGTCTGCAGGTC
GACTCCAGAAAACATGAGGATCACCCATGTCTGCAGTATTCCCGGGTTCATTAGATCTGC
GCGCGATCGATATCAGCGCTTTAAATTTGCG**CTCGAG**

Fig 13. Schematic features of the construction of the three major editing factors. Sequence result of the guideRNA-MS2-6X, Here Bold underline are the restriction sites. In the upstream EcoRI and downstream XhoI. The bold part is the 19bp guideRNA and rest are the MS2-6X part

2.2.4 Preparation of Double MS2 (1X MS2 on either side of the guideRNA)

For the preparation of the 1X MS2 on the either side of the guideRNA, four oligos were designed, which was done by following the idea of Katrekar et al., 2019 (27). Then the annealing and kination of the oligos was done. Oligo 1+Oligo 2 and Oligo 3+Oligo 4 was annealed and kinated (T4 polynucleotide kinase, 10XC kinase buffer, Biolab). The pol III U6 promoter containing plasmid was digested with the BbsI (New England Biolabs) restriction enzyme. BAP treatment was done to the digested plasmid. Finally three- way ligation was done among the two annealed products and one digested vector. After that the ligated product was transformed into DH5∝ competent cells and positive colonies were collected. Plasmid DNA was extracted by using the Easy pure plasmid extraction kit. For the confirmation of the construct the sequencing of the sample was done.

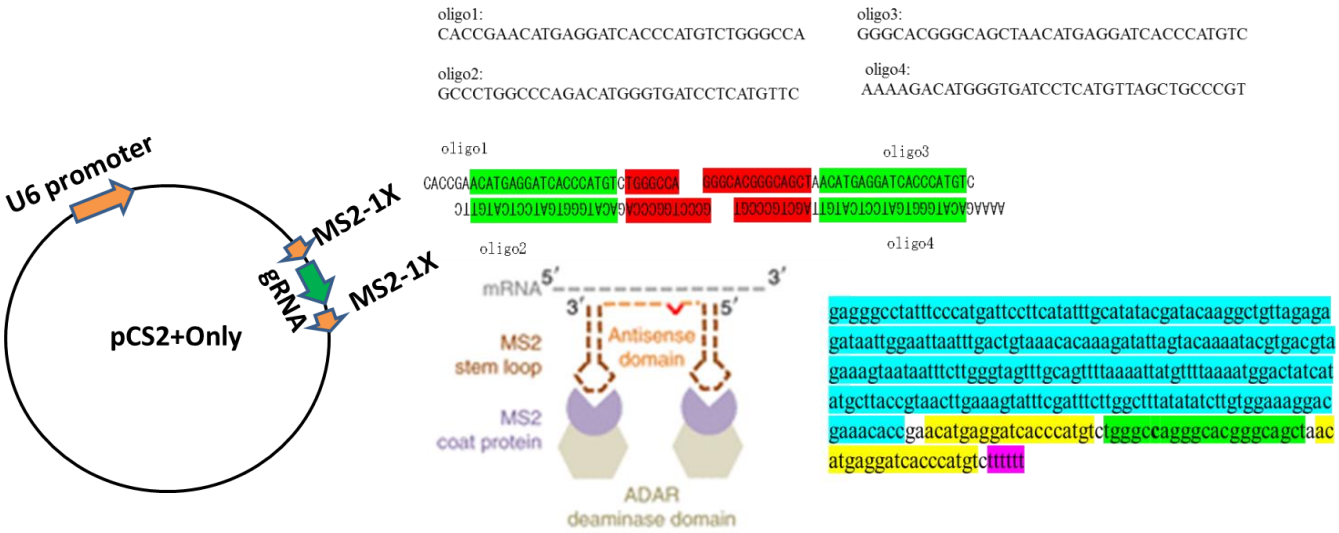


Fig. 14 Preparation of the MS2-1X stem loop (Double) on the either side of the 21 bp guideRNA in the pCS2+only plasmid vector under the control of U6 promoter

2.2.5 Cell culture:

Human embryonic kidney (HEK) 293 cells were cultured in 60 mm glass bottom dishes in high-glucose Dulbecco's Modified Eagle's medium (D-MEM, Gibco), supplemented with 10% fetal bovine serum (FBS, Gibco), at 37°C and in an atmosphere containing 5% CO_2 for 24 hours until they reached 80–100% confluence.

2.2.6 Transfection with Lipofectamine 2000:

HEK293 cells were transfected with plasmid using Lipofectamine 2000 (Invitrogen), according to the manufacturer's instructions, with 3×10^5 cells cultured per well of a 24-well plate. After 24 hours, when the cells were 50–70% confluent, transfection was assessed by microscopy. Cells were transfected with all three factors using Optimem (Gibco) and Lipofectamine 2000 (Invitrogen), according to the manufacturers' instructions. After 6 hours, Optimem was replaced by D-MEM (Gibco) to optimize cell growth, and the cells were incubated for 48 hours.

2.2.7 Cell expression of GFP:

GFP expression was monitored by JuLi fluorescence microscopy, to assess the appearance of green fluorescence, and LSM confocal fluorescence microscopy, to assess cellular localization of fluorescence. The experiment of JuLi and confocal was done independently so that transfection efficiencies should be different.

2.2.8 RNA extraction and cDNA synthesis:

Positive cells for restoration having expression were harvested by the filter paper after observing under microscope and pointing the particular area, their total RNA was extracted using Trizol reagent (Invitrogen) method. cDNA was synthesized from extracted RNA using a SuperScript™ III First-Strand Synthesis System for RT-PCR

(Invitrogen). During the collection of the cells there were also some non-expressed cells as it is almost impossible to harvest only the positive cells having the expression as I did not use any cell sorting technique.

2.2.9 PCR-RFLP:

To confirm restoration of EGFP expression, cDNA was PCR amplified using Gotaq polymerase (Promega), and the PCR product was digested with the restriction enzyme BmgT120I (Takara). The primers for the PCR-RFLP are Forward primer: ACGTAAACGGCCACAAGTTC, Reverse primer: CTTGTAGTTGCCGTCGTCCT The intensity of the representing unedited and edited mRNA was measured using ImageJ software, and the ratio of intensity for each sample was calculated.

2.10 Direct sequencing by Sanger's method:

PCR amplified fragments were electrophoresed on 6% polyacrylamide gels. The bands were cut, and the DNA was extracted with 0.1 X TE. A 2 µl aliquot of each sample was subjected to sequencing using a Big Dye Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Technologies), according to the manufacturer's instructions. The rate of editing efficiency was calculated by measuring the area and peak height using ImageJ software (JAVA).

Editing rates were calculated from peak areas using the following equations:

$$\text{Editing rate (for sense)} = \frac{\text{Area of 5' G or 3'G}}{\text{Area of 5'C or 3'C + Area of 5'T or 3'T}} \text{ X } 100$$

$$\text{Editing rate (for antisense)} = \frac{\text{Area of 5' C or 3'C}}{\text{Area of 5' G or 3'G + Area of 5'A or}} \text{ X } 100$$

Editing rates were calculated from peak heights using the following equations:

$$\text{Editing rate (for sense)} = \frac{\text{Height of 5'G or 3'G}}{\text{Height of 5' G or 3'G + Height of 5'A or 3'A}} \text{ X } 100$$

$$\text{Editing rate (for antisense)} = \frac{\text{Height of 5'C or 3'C}}{\text{Height of 5' C or 3'C + Height of 5'T or}}$$

Each experiment was repeated for sense three times and for antisense twice, and their means were calculated.

2.3 RESULTS

2.3.1 Microscopic observation:

HEK cells transfected with mutated EGFP containing the TAA stop codon showed no evidence of fluorescence signals (Figure 15, 16), indicating that the mutation in amino acid 58 EGFP, TGG (Trp) to TAA (stop codon), completely abolished fluorescence expression. Cells transfected with mutated EGFP containing this stop codon, alone or in combination with plasmids encoding either ADAR1 or guide RNA, were also negative for fluorescence (Figure 15a-d). By contrast, cells transfected with all three plasmids were positive for fluorescence expression (Figure 15e), as were cells transfected with vector expressing wild-type EGFP (Figure 15f), this part has already been presented in Azad et al., 2017. Because JuLi fluorescence microscopy was unable to determine the exact location of fluorescence, it was unclear whether fluorescence derived from these cells was an artifact. LSM confocal microscopy, however, showed fluorescence within the cells transfected with all three factors (ochre-EGFP+ADAR1+guide RNA), but not with any combination of two factors (Figure 16a-c). LSM confocal microscopy also showed that the restored EGFP was localized to part of the cytoplasm, whereas wild-type EGFP was expressed throughout the cell.

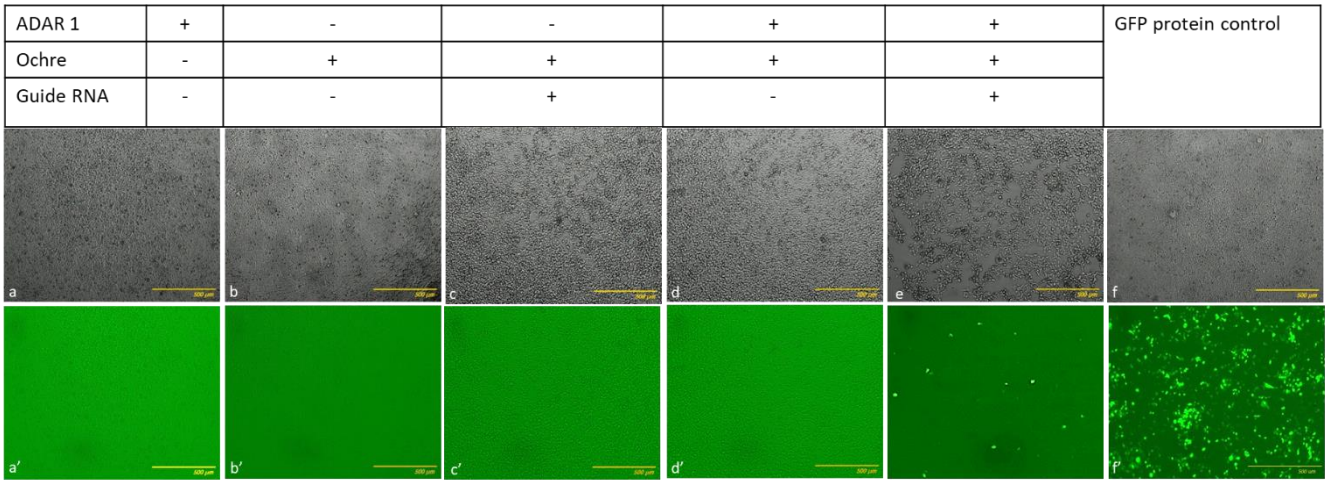


Fig 15. Transfection in HEK cell with wild type EGFP, 1 factor (Only EGFP containing ochre stop codon or Only ADAR1), 2 factors (EGFP containing Ochre stop codon + ADAR1 or EGFP containing Ochre stop codon + Guide RNA) and 3 factors (EGFP containing Ochre stop codon + ADAR1 + Guide RNA). Only the wild type EGFP and the 3 factors (EGFP containing Ochre stop codon +ADAR1+Guide RNA) express the fluorescence expression due to the presence of TGG in wild type and editing from TAA to TGG by the 3 factors (EGFP containing Ochre stop codon +ADAR1+Guide RNA) Imaging by JuLi smart fluorescence microscope.

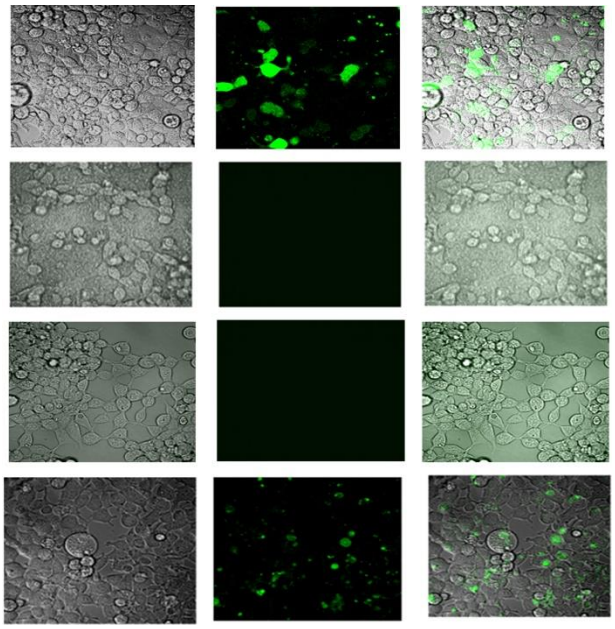


Fig 16. Transfection in HEK cell with wild type EGFP, 2 factors (Ochre+ADAR1) or (Ochre+gRNA) and 3 factors (Ochre+ADAR1+Guide RNA). Only the wild type EGFP (**a**) and the 3 factors (Oche+ADAR1+Guide RNA) (d) express the fluorescence expression due to the presence of TGG in wild type and editing from TAA to TGG by the 3 factors (Oche+ADAR1+Guide RNA). However, the image of two factors (b & c) does not show any fluorescence expression as the TAA has not been restored by the two factors. *Ochre: EGFP containing ochre stop codon

2.3.2 Confirmation of the restoration by PCR-RFLP:

RT-PCR followed by RFLP further confirmed restoration of coding by the TAA stop codon. The restriction enzyme used, BmgT120I, could digest both wild EGFP (5'…GGNCC…3') and restored EGFP into bands of 160 bp and 100 bp, but could not digest unrestored EGFP, which remained as a single band of 260 bp (Figure 4). RT-PCR-RFLP assays showed that RNA extracted from cells transfected with plasmids encoding Mutated EGFP-ochre+ADAR1 or Mutated EGFP-ochre+guide RNA) remained uncut. By comparison, 13% of the RNA extracted from cells transfected with plasmids expressing Mutated EGFP-ochre+ADAR1+guide RNA remained uncut, whereas 87% was digested by BmgT120I (Table 3).

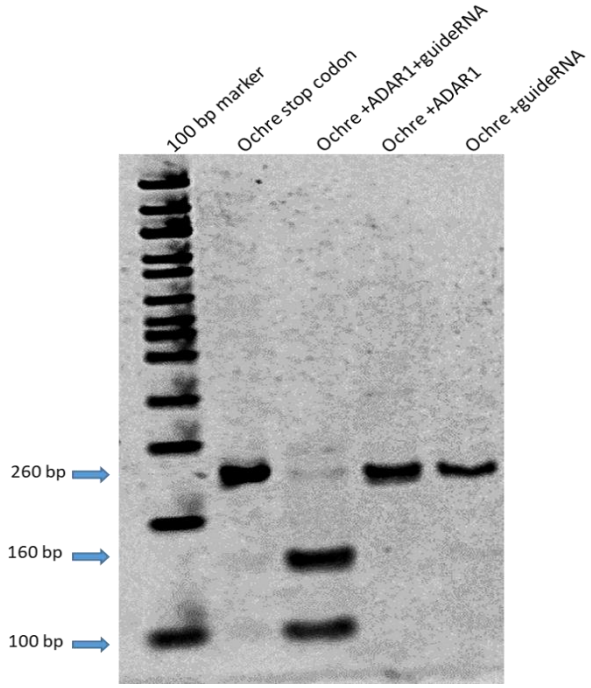


Fig 17. Representation of both the PCR and RFLP results, cut (wild type and restored) and uncut (single factor), after PCR-RFLP using BmgT120I restriction enzyme. For cut the band was found at 160 base pair and 100 base pair, whereas the uncut remain at 260 base pair as the only PCR result

Table 3. Ratio of the edited and unedited from PCR-RFLP

	Area	After subtraction	Mean	Mean Total	Std Dev	Std Error	Ratio of Uncut	Percentage of edited part
Subtract background	2.75							
					0.28	0.195		
Uncut	4.14	1.39	1.19+/-0.195					(100-13)
	3.75	1						
					2.73	1.93		87%
Cut 160 bp	6.13	3.38	5.31+/-1.93	9.11			0.13 13%	
	9.99	7.24						
					1.12	0.795		
Cut 100 bp	6.15	3.4	2.60+/-0.795					
	4.56	1.81						

2.3.3 Direct sequencing of restored EGFP mRNA:

The restored plasmid was directly sequenced using both the sense and antisense primers. cDNA extracted from cells transfected with ochre-EGFP+ADAR1+guide RNA showed a double peak, representing restored G and mutated A nucleotides (Figure 5a, 5b). Sequencing using both sense and antisense primers confirmed that editing was successful, whereas none of the surrounding A residues (19 bp upstream and downstream from the target) was altered. It means there was no off-target. The peak height of the 5'G was greater than that of the 3'G, indicating that editing was greater for the 5'A than for the 3'A. Moreover, peak quality was better for antisense than for sense sequencing. The results obtained from all cells, those with and without fluorescence, were very similar to those obtained for fluorescent-positive cells alone. The results were also shown to be reproducible over three different experiments (Table 4a, 4b and 5a, 5b).

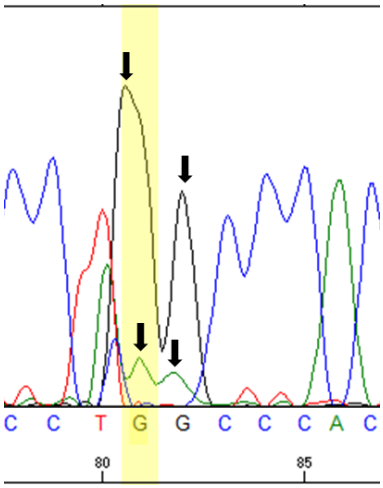


Fig 18a. Restored EGFP from ochre stop codon (TAA) to Normal codon TGG (Sense primer).

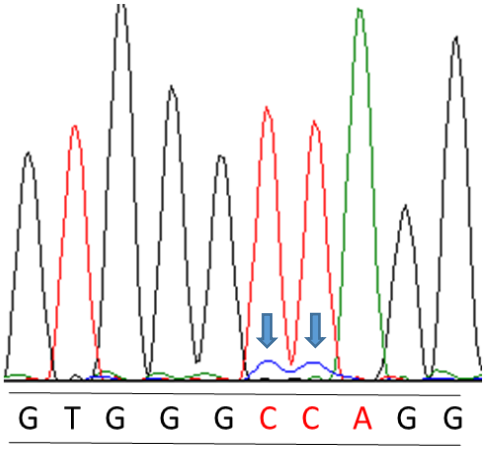


Fig 18b. Restored EGFP from ochre stop codon (TAA) to Normal codon TGG (Anti-sense primer), Showing the restoration of the TAA to TGG.

Table 4a. Calculation of editing efficiency considering area by using sense primer

	Mean height	Std Dev	Std Error	Mean Width	Std Dev	Std Error	Mean Area	Editing rate
5'G	271.33+/-32.41	56.13	32.41	23+/-2.52	4.36	2.52	6240.59	0.95* 95%
3'G	128+/-16.37	28.35	16.37	18.33+/-0.88	1.53	0.88	2346.24	0.91 91%
5'A	21+/-5.51	9.54	5.51	16.33+/-1.76	3.06	1.76	342.93	
3'A	17.67+/-2.60	4.51	2.60	13.33+/-3.76	6.51	3.7	235.54	

**P value is less than 0.05 thus null hypothesis is significantly true means 5'A is more edited than 3'A

Table 4b. Calculation of editing efficiency considering area by using antisense primer

	Mean Height	Std Dev	Std Error	Editing rate
5' G	271.33+/-32.41	56.13	32.41	0.93* 93%
3'G	128+/-16.37	28.35	16.37	0.88 88%
5'A	21+/-5.51	9.54	5.51	
3'A	17.67+/-2.60	4.51	2.60	

**P value is less than 0.05 thus null hypothesis is significantly true means 5'A is more edited than 3'A

Table 5a. Calculation of editing efficiency considering peak height by using Sense primer

	Mean height	Std Dev	Std Error	Mean Width	Std Dev	Std Error	Mean Area	Editing Rat
3' G	222.5 +/-1.5	2.12	1.5	16+/-0	0	0	3560	0.89 89%
5' G	210.5+/-1.5	2.12	1.5	16+/-0	0	0	3368	0.91* 91%
3' A	21+/-0	0	0	21+/-1.0	1.41	1	441	
5' A	19+/-3	4.24	3	17.5+/-0.5	0.71	0.5	332.5	

**P value is less than 0.05 thus null hypothesis is significantly true means 5'A is more edited than 3'A

Table 5b. Calculation of editing efficiency considering peak height by using antisense primer

	Mean Height	Std Dev	Std Error	Editing Rate
3' G	222.5+/-1.5	2.12	1.5	0.91 91%
5' G	210.5+/-1.5	2.12	1.5	0.92* 92%
3' A	21+/-0.0	0	0	
5' A	19+/-0.0	4.24	3	

**P value is less than 0.05 thus null hypothesis is significantly true means 5'A is more edited than 3'A

2.4 Application of the double MS2 on the either side of the guide with ADAR1 deaminase

2.4.1 Juli Microscopic observation for confirmation of the restoration:

HEK 293 cells transfected with mutated EGFP (TAA, Ochre stop codon) showed no evidence of fluorescence signals (Figure 19), indicating that the mutation in 58th amino acid, TGG (Trp) to TAA (stop codon), completely abolished fluorescence expression. Cells transfected with mutated EGFP containing this stop codon, alone or in combination with plasmids encoding either ADAR1 or guide RNA, were also negative for fluorescence (Figure. 19a-d). By contrast, cells transfected with all three plasmids were positive for fluorescence expression (Figure 19e), as were cells transfected with vector expressing wild-type EGFP (Figure 19f). So only one factor either deaminase or guide do not possess any capacity to restore the genetic code.

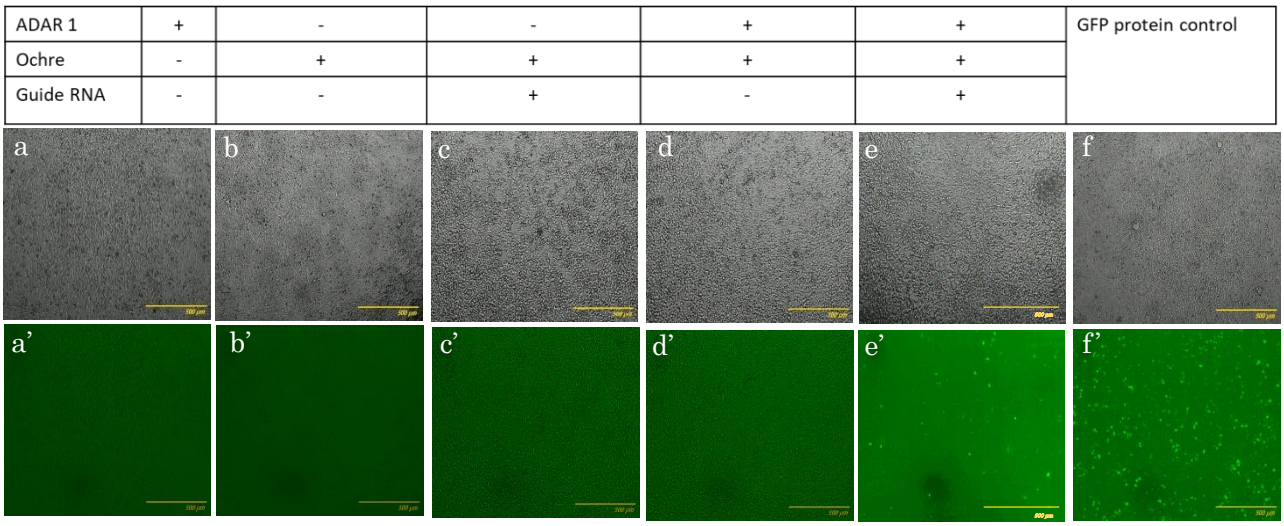


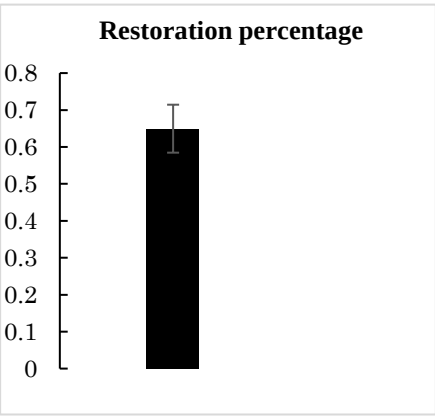
Fig 19. Transfection in HEK 293 cell with wild type EGFP, 1 factor (Only EGFP containing Ochre stop codon or Only ADAR1), 2 factors (EGFP containing Ochre stop codon + ADAR1 or EGFP containing Ochre stop codon + Guide RNA) and 3 factors (EGFP containing Ochre stop codon + ADAR1 + Guide RNA). Only the wild type EGFP and the 3 factors (EGFP containing Ochre stop codon +ADAR1+Guide RNA) express the fluorescence expression due to the presence of TGG in wild type and editing from TAA to TGG by the 3 factors (EGFP containing Ochre stop codon +ADAR1+Guide RNA) Imaging by JuLi smart fluorescence microscope.

2.4.2 PCR-RFLP for confirmation of the restoration:

RT-PCR followed by RFLP further confirmed restoration of coding by the TAA stop codon. The restriction enzyme used, BmgT120I, could digest both wild EGFP (5'…GGNCC…3') and restored EGFP into bands of 160 bp and 100 bp, but could not digest unrestored EGFP, which remained as a single band of 260 bp (Figure 20). RT-PCR-RFLP assays showed that RNA extracted from cells transfected with plasmids encoding Mutated EGFP-ochre+ADAR1 or Mutated EGFP-ochre+guide RNA) remained uncut. By comparison, 35% of the RNA extracted from cells transfected with plasmids expressing Mutated EGFP-ochre+ADAR1+guide RNA remained uncut, whereas 65% was digested by BmgT120I.

Substrat 2.75

		After substract		STD Dev.	STD error	Restoration percentage	
	4.53	1.78					
Uncut band	4.44	1.69	1.63	0.182	0.10	0.65	65%
at 260 bp	4.18	1.43					
	4.22	1.47					
Cut band	4.21	1.46	1.5	0.061	0.04		
at 150 bp	4.32	1.57					
	4.23	1.48					
Cut band	4.17	1.42	1.53	0.14	0.08		
at 110 bp	4.43	1.68					



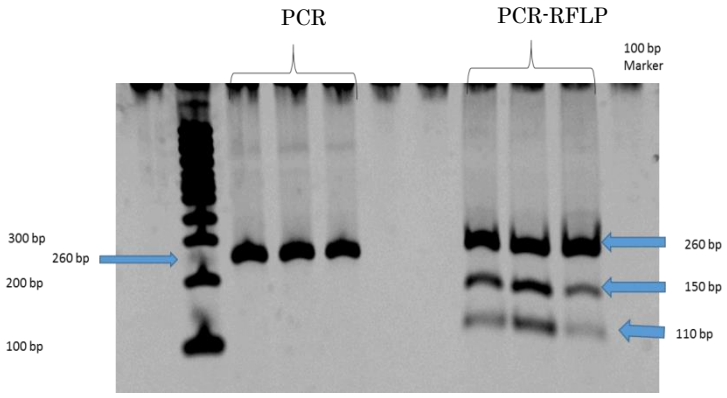
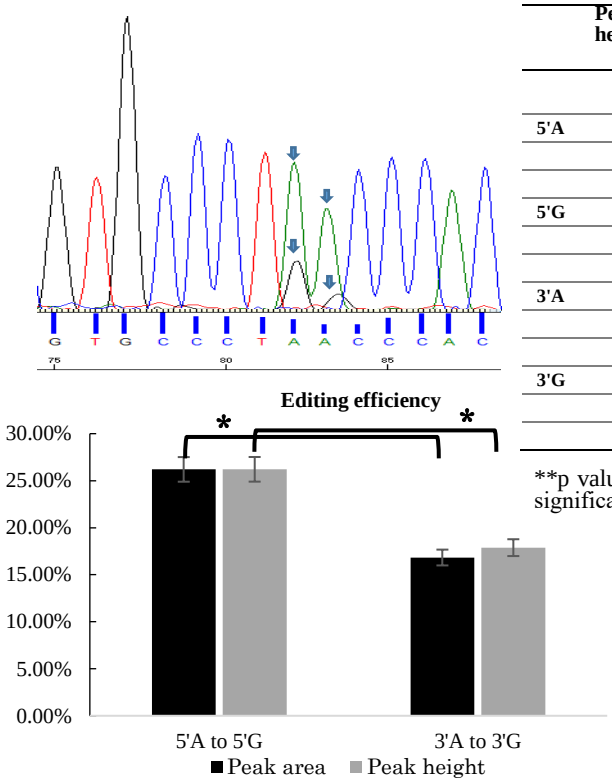


Fig 20. Representation of both the PCR and RFLP results, cut (wild type and restored) and uncut (single factor), after PCR-RFLP using BmgT120I restriction enzyme. For cut the band was found at 160 base pair and 100 base pair, whereas the uncut remain at 260 base pair as the only PCR result also showed that the amplified DNA was at 260 bp length, For all data the statistical analysis (mean±SEM) was done, where (n=3)

2.4.3 Direct sequencing of restored EGFP mRNA:

The restored plasmid was directly sequenced using both the sense primer. cDNA extracted from cells transfected with ochre-EGFP+ADAR1+guide RNA showed a double peak, representing restored G and mutated A nucleotides (Figure 21). Sequencing using sense primer confirmed that editing was successful, whereas none of the surrounding A residues (21 bp upstream and downstream from the target) was altered. It means there was no off-target. The peak height of the 5'G was greater than that of the 3'G, indicating that editing was greater for the 5'A than for the 3'A. The results obtained from all cells, those with and without fluorescence, were very similar to those obtained for fluorescent-positive cells alone. The results were also shown to be reproducible over three different experiments.



	Peak height	Peak width	Avg peak height	Avg peak width	Peak area	Avg peak area	Std Dev (SD)	Std error(SE)		
5'A	212	17	209.33	17	3604	3558.67	7.37			4
	201	18			3618					
	215	16			3440					
5'G	70	16	74.33	17	1120	1263.67	4.04			2
	78	18			1404					
	75	17			1275					
3'A	147		22	141.67		22	3234	3116.67	5.51	3.18
	136		21				2856			
	142		23				3266			
3'G	27		25	28.67		23.67	675	678.44	1.53	0.88
	30		23				690			
	29		23				667			

**p value is less than 0.01, it means the null hypothesis is true which means the 5'A is significantly more edited than 3'A

Fig 21. Restored EGFP from ochre stop codon (TAA) to Normal codon TGG, the double peak is there and the 5'A is more edited than the 3'. According to the editing rate calculation, 26% of 5'A and 17% of 3'A was restored from TAA to TGG by peak area and peak height, respectively. For all data the statistical analysis (mean±SEM) was done, where (n=3).

2.5 DISCUSSION

This study described the establishment of a system, involving an artificial deaminase system and MS2, to restore a coding sequence to a mutated ochre stop codon TAA in the protein EGFP. ADAR1 can be directed to editing RNA with the MS2 system. The guide RNA was simple, and each such guide RNA can be designed for specific targets in other cDNAs along with the MS2 system. ADAR1 can edit the stop codon TAA, yielding a readable codon TGG, as shown by JuLi smart fluorescence microscopy, LSM confocal microscopy, PCR-RFLP, and direct sequencing. The ochre stop codon TAA is more difficult to restore than that of the other stop codons, as TAA requires changes in two nucleotides. ADAR2 lacking a double-stranded binding domain was shown to be functional, converting A to I for RNA editing in vitro and in vivo (13). In addition, the Lambda N system has been utilized to control site directed RNA editing. The MS2 system is more frequently used than the Lambda N system in studies of RNA. Chemical modification of guide RNA allows its attachment to the deaminase domains of ADAR1 and ADAR2, with the resulting constructs having deaminase activity in vitro, in cell culture, and in annelid worms (5, 18, 23). Transfection of cells with all three factors resulted in the conversion of truncated to full-length EGFP. EGFP signals, however, were not detected following transfection of mutated EGFP alone or mutated EGFP and guide RNA treatments, indicating that other cellular factors are not involved in editing. Further tuning of this system is required, including optimization of guide RNA, rotating the orientation of the MS2 system to the enzyme, and optimizing the ratios of the three factors. Optimizing the concentrations of enzyme and reporter substrate can eliminate minor auto editing (5).

PCR-RFLP was performed to determine whether the specified nucleotide bases were

present in the targeted sequence after editing. The enzyme used in this study, BmgT120I, is specific for the sequence 5'--G | GNCC--3'. cDNA from cells transfected with ochre-Mutated EGFP, alone or with ADAR1 or guide RNA, remained uncut, yielding a single 260 bp band. By contrast, cDNA from cells transfected with wild-type EGFP or with ochre-Mutated EGFP + ADAR1 + guide RNA were digested by BmgT120I, yielding two bands at 160 bp and 100 bp, indicating that this system restored TAA to TGG. Interestingly, however, a minor uncut band of 260 bp was observed in cDNA from cells transfected with the three factors, indicating that most (87%) but not all of the TAA was converted to TGG, with 13% remaining unconverted.

Direct sequencing using the Sanger method showed that the mutated sequence with the ochre stop codon was CCTAACC, whereas the restored sequences were CCTGGCC with the sense primer and GGCCAGG with the antisense primer. Editing rates were calculated based on both peak areas and peak heights for both the sense and antisense primers. During the editing process, TAA may be first converted to TAG and subsequently to TGG. Using both sense and antisense primers, the rate of editing was higher for the 5'A than for the 3'A (10). Editing rates using the sense primer were 95% for the 5'A and 91% for the 3'A, based on peak area, and 86% and 85%, respectively, based on peak height. By comparison, editing rates using the antisense primer were 91% for the 5'A and 89% for the 3'A, based on peak area, and 92% and 91%, respectively, based on peak height. When all cells, not only fluorescent-positive cells, were included, the editing rates for the 5'A and 3'A were 7.05% and 6.66%, respectively. These results were not in conflict with the microscopy and PCR-RFLP results. In case of the double mutation there is a possibility that, during the process of Editing the TAA will be first converted or edited to TAG and later on the TGG will be found. It means the 3' A will

be edited fast and later on the 5'. But in both the sense and antisense the rate of editing is higher in case of the 5' A than that of the 3' A (10).

Later on I applied the double MS2 stem loop guide RNA along with the ADAR1 deaminase into the ochre stop codon. From the results it was found that after application of both the editing factors the restoration of the genetic code from TAA to TGG happened. The Juli microscopic observation evident the fact. The PCR-RFLP results also showed that after the restriction digestion the two cut bands were found at 160 and 100 bp. According to the sequence preference of the restriction enzyme the wild type GFP and restored GFP should cut and the result also support the data.

From the Sanger's sequencing result I could see the double peaks in case of the targeted 5'A and 3'A after the application of both the editing factors (deaminase and guideRNA). From the calculation of the editing rate it was found that 5'A is edited approx. 26% according to peak height and peak area, whereas in case of the 3'A it is approx. 17%. In this experiment the whole dish of the cells were collected for cDNA synthesis. Here, the p value was 0.002 which is less than 0.01 at 95% confidence interval which proved that 5'A is significantly more restored than that of the 3'A.

After application of the double MS2 containing guideRNA the editing rate has been increased compared to the 6X MS2 stem loop downstream to the guideRNA. The experimental findings were also supported by Katrekar et al., (2019), who performed the experiment *in vivo*, on mdx mouse and found this concept of preparation of the guideRNA worked better than the 6X MS2-stem loop and also they observed after addition of a NES with the guide would work even better (27). Which could be applied in my research work in the near future.

2.6 CONCLUSION

The findings of this study suggested that this developed artificial deaminase system successfully restored the gene expression by editing point mutations, which might be applied in the RNA of patients with various diseases and may even mitigate symptoms of patients. This system altered A-to-I (G), with greater restoration of the 5'A than the 3'A in the sequence of UAA. The ability of the MS2 system with ADAR1 deaminase domain and guide RNA to edit the mutated RNA sequences may overcome genetic diseases caused by the introduction of stop codons. Future studies should include the optimization of editing efficiency, by using the guide RNAs of different lengths and altering the MS2 stem loop, such as we have utilized the 1XMS2 (double MS2) stem loop on either side of guideRNA could be modified to 2X or 3X on either side. If possible, the practical application of this system should be tested in animal models.

2.6 REFERENCES

1. Araki M., and Ishii T. Providing Appropriate Risk Information on Genome Editing for Patients. Trends Biotechno, 134(2): 86–90, (2016).
2. 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).
3. Bertrand E., Chartrand P., Schaefer M., Shenoy S.M., Singer R.H., and Long R.M. Localization of ASH1 mRNA particles in living yeast. Molecular Cell, 2(4): 437–45, (1998).
4. Eggington J.M., Greene T. and Bass B.L. Nat. Commun., 2: 319, (2011). 10. 1038/ncomms1324
5. 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).
6. Heidenreich M., and Zhang F. Applications of CRISPR-Cas systems in neuroscience. Nat Rev Neurosci., 17(1): 36–44, (2016).
7. Hsu P.D., Lander E.S., and Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. Cell, 157(6): 1262–78, (2014).
8. Jacob C. How to make a Guide RNA for a Cas9 knockout. Innovative genomics Institute (2014).
9. Johansson H.E., Liljas L., and Uhlenbeck O.C. RNA recognition by the MS2 phage coat protein. Sem Virol, 8(3): 176–185, (1997).
10. Katrekar D., and Mali P. In vivo RNA targeting of point mutations via suppressor tRNAs and adenosine deaminases. Cold Spring Harbour (BioRxiv), (2017).

11. Kim Y.G., Cha J., and Chandrasegaran S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proceedings of the National Academy of Sciences, 93(3): 1156–1160, (1996).
12. Mali P., Luhan Y., Kevin M.E., John A., Marc G., James E.D., et al. RNA-guided human genome engineering via Cas9, Science 339(6121): 823–826, (2013).
13. 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).
14. Nishikura K. Functions and regulation of RNA editing by ADAR deaminases. Annu. Rev. Biochem, 79, 321–349, (2010).
15. Porteus M.H., and Carroll D. Gene targeting using zinc finger nucleases. Nat Biotechnol, 23(8): 967–973, (2005).
16. Renaud J.B., Boix C., Charpentier M., De-Cian A., Cochennec J., Duvernois-Berthet E., Perrouault L., Tesson L., Edouard J., Thinard R., Cherifi Y., Menoret S., Fontanière S., De-Crozé N., Fraichard A., Sohm F., Anegon I.,Concordet J.P. and Giovannangeli C. Cell Rep., 14: 2263–2272, (2016).
17. Rinkevich F.D., Schweitzer P.A. and Scott J.G. BMC Res. Notes, 5:1–6, (2015).
18. Schneider M.F., Wettengel J., Hoffmann P.C., and Stafforst T. Optimal guideRNAs for re-directing deaminase activity of hADAR1 and hADAR2 in trans. Nucl. Acids Res, 42(10): e87, (2004).
19. Stafforst T., and Schneider M.F. An RNA Deaminase conjugates electively repairs point mutations. Angew. Chem. Int. Ed, 51(44): 11166–11169, (2012).
20. Thomas G., Shannon J.S., Sai-lan S., and Jia L. Genome editing technolo-gies:

principles and applications. Cold spring Harbor Perspectives in Biology, (2016).

21. 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).
22. 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).
23. Wettengel J., Reautschnig P., Geisler S., Kahle P.J., and Stafforst T. Harnessing human ADAR2 for 418 RNA repair–recoding a PINK1 mutation rescues mitophagy. Nucleic Acids Res, 45(5): 2797-2808, (2017).
24. Zhang F., Le C., Simona L., Sriram K., George M.C., and Paola A. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. Nat Biotechnol, 29(2): 149–153, (2011).
25. 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).
26. Bhakta S., Azad M.T.A. and Tsukahara T. Genetic code restoration by artificial RNA editing of Ochre stop codon with ADAR1deaminase. Protein Engineering, Design & Selection, 31(12): 1-8, (2018).
27. Katrekar D., Chen G., Meluzzi D., Ganesh A., Worlikar A., Shih Y., Varghese S. and Mali P. In vivo RNA editing of point mutations via RNAguided adenosine deaminases. Nature Methods, 16: 239–242, (2019).
28. Bhakta S., and Tsukahara T. Artificial RNA Editing with ADAR for Gene Therapy. Current Gene therapy, 20 (1): 44-54, (2020).

CHAPTER III

Chapter III: Genetic code restoration in BFP (derivative of GFP) by artificial RNA editing using APOBEC 1 cytidine deaminase

3.1 INTRODUCTION

Genetic engineering, a set of technologies for regulating the expression and functions of intracellular targeted genes, has been broadly utilized in fundamental research as well as medicinal and therapeutic applications (1). Nowadays, several other genome editing techniques have made it possible to manipulate genomic information in a targeted manner (1, 2, 3 4, 5). These methods include zinc-finger nucleases (ZFNs), transcription activator–like effector (TALE) nucleases (TALENs), and clustered regularly interspaced short palindromic repeats (CRISPR)–CRISPR-associated protein 9 (Cas9) (6, 7). These enzymes can be introduced into cells for a technique known as genome editing with engineered nucleases. The latest and currently popular enzyme for the genome editing is CRISPR/ Cas9. The CRISPR/Cas9 framework originally developed as an adaptive immune system that defends bacteria and archaea from viral (phage) and plasmid infection. But in recent years, CRISPR/Cas9 has been adapted for genome editing, representing an important scientific breakthrough. CRISPR/Cas9 can mediate gene modifications at specific locations, allowing scientists to rapidly and efficiently edit the genomes of a wide range of organisms. Accordingly, this system also has the potential to be used in cell therapy. The CRISPR/Cas9 system consists of several components of which the Cas9 protein itself and the sgRNA (synthetic single guideRNA) are essential for genome editing (47), which is also considered to be very effective system for controlling off target effect during genome editing (43).

Base editing is another type of genome editing that enables direct, irreversible transformation of one base pair to another at a target genomic locus without requiring double-stranded DNA breaks, the most commonly utilized base editors are third-generation designs (BE3) consisting of a catalytically impaired CRISPR–Cas9

mutant (41, 48). On the other hand, these existing base editors also sometimes create undesired C-to-T modifications when more than one C is present in the enzyme's five-base-pair editing window which the Gehrke et al., group has attempted to reduce bystander mutations using an engineered human APOBEC3A (eA3A) domain with Cas9 (46).

Although these techniques are expected to find applications in the treatment of diseases, it remains difficult to achieve accurate genome editing in all affected cells. Moreover, incorrect genome editing has the potential to cause cancers and other diseases.

Therefore, I think genome editing is a suitable method for treatment of fertilized eggs or *ex vivo*. To treat patients, it is needed to deliver the gene editing tools to trillions of cells and various tissues, so genome editing is not a suitable technique for gene therapy for the treatment of patients. By contrast, RNA is expressed in all tissues and cells, and if RNA repair errors occur due to incorrect RNA editing, the mutated RNAs are quickly degraded and do not affect the genome sequence (8, 9). Therefore, from the standpoint of patient treatment, RNA editing is preferable than genome editing. Accordingly, I have developed an artificial RNA editing system, based on the APOBEC1 deaminase, for restoration of the wild-type genetic code in genetic diseases caused by T-to-C mutations. Some examples of such disorders include: ADA deficiency, cystic fibrosis, elliptocytosis, antithrombin III deficiency, and others (10, 11, 12, 13). A search of the NCBI disease mutation database ClinVar revealed 762 registered cases of T-to-C mutations (Supplementary Table S1).

In previous studies, artificial A-to-I RNA editing was done by using ADAR family enzymes tethered to gRNAs complementary to the target sequence (14, 39). For tethering, Stafforst and colleagues used the SNAP-tag (15, 16, 17), Montiel-Gonzalez and colleagues used Lambda-N protein and box B element RNA (18), and our group

used the MS2 system (19, 20).

To date, however, no study has been published involving MS2 system for C-to-U RNA editing. I postulated that an artificial RNA editase for C-to-U conversion could be designed similarly to the A-to-I editases reported previously. For the purpose apolipoprotein B-mRNA editing family member was used.

Members of the apolipoprotein B mRNA editing enzyme, catalytic polypeptide (APOBEC) and activation-induced cytidine deaminase (AICDA/AID) families are single strand explicit cytidine deaminases, expressed in numerous cells and tissues, which catalyze cytidine-to-uridine (C-to-U) base substitutions in RNA, viral DNA, and genomic DNA (21). APOBEC1, a member of the APOBEC family can also perform editing on its own *in vitro*, this has been confirmed by overexpression in cell lines, as well as *in vivo* in the absence of cofactors (22, 23). Previously, it was thought that cis-acting components and trans-acting factors are essential and adequate to carry out C-to-U RNA editing *in vitro*. Consistent with this, one group of researchers reported that pure GST-APOBEC1 can edit apoB mRNA (having an optimal amount) without any auxiliary factors, moreover, this apo-enzyme is temperature stable (24).

Zenpei and colleagues (2017) (25), reported a combination of CRISPR-Cas9 and activation-induced cytidine deaminase (Target-AID) for site-directed mutagenesis at genomic regions specified by single guide RNAs (sgRNAs); their system could be used for genetic engineering in harvesting plants such as rice and tomato (25). Mali and colleagues (2013) focused on human genome editing with the utilization of the Cas9 (26). But these studies were done by considering the genome editing not RNA editing, which is our concern. So in this study, I have focused on an artificial C-to-U RNA editing system as a tool for *de novo* gene therapy.

Using an MS2-tagged system, I previously restored the original sequence of a G-to-A

mutation via A-to-I editing with the ADAR1 deaminase (19, 20, 27). Here, by contrast, I sought to restore C-to-U in the context of a T-to-C mutation using our artificial enzyme system along with a specific gRNA. Tagging with MS2 is based on the natural interaction between the MS2 bacteriophage coat protein and a stem-loop structure from the phage genome (28), which has been used for biochemical purification of RNA–protein complexes and combined with green fluorescent protein (GFP) expression to enable detection of RNA in living cells (29). By using the phenomena in bacteriophage regarding the coat protein and stem loop, APOBEC 1 was bound to the MS2 coat protein and the gRNA bound to MS2 stem loop with a view to initiate the binding of the coat protein and stem loop thus allowing the gRNA to guide the deaminase at the specific location/site and perform editing at the targeted nucleotide sequence. This is the first study in which the MS2 system has been engineered to create an RNA editing enzyme complex that is capable of targeting a specific point mutation. Using the GFP point mutant blue fluorescence protein (BFP) as a model target RNA, our artificial RNA editing system could successfully convert C-to-U at the mRNA level, restoring the wild-type sequence.

3.2 MATERIALS AND METHODS

3.2.1 Target plasmid (mutated EGFP or BFP) construct preparation:

The target BFP gene was prepared by a point mutation at 199th position of GFP by following the protocols of Luyen *et al.*, (2012, 2015) (33, 35). The BFP construct was transfected into HEK 293 cells and were selected with G418 @ 500 ng/mL for few weeks. Positive colonies were picked and confirmed by Sanger's sequencing.

3.2.2 APOBEC1 deaminase plasmid construction:

To target the enzyme to the BFP was chosen as a codon of interest; I cloned the deaminase domain of APOBEC1 downstream of MS2 in pCS2+MT vector under the control of the pol II CMV IE-94 promoter, using the *Xho*I and *Xba*I (Takara, Shiga, Japan) restriction sites. The resultant plasmid was designated as pCS2+MT-MS2HB-APOBEC1. APOBEC1 was PCR-amplified from HEK 293 cell line using primers with the appropriate restriction sites: *Xho*I catalytic APOBEC1 Fw: tccactcgagatgccctgggagtttgacgtctt; *Xba*I catalytic APOBEC1 Rv: acggtctagattaagggtgccgactcagaaactc; (red color font indicates restriction sites, blue color font highlighting the tri-nucleotide atg/tta is a leader sequence allowing proper recognition by the restriction enzyme). Positive colonies were picked and confirmed by Sanger sequencing. The open reading frame and catalytic domain were confirmed using the ExPASY Bioinformatics resource portal and NCBI-BLAST.

3.2.3 Preparation of the gRNA to direct the deaminase to target:

To prepare the guideRNA, a 21 ntd sequence complementary sequence to the target mRNA having a mismatch A at the target C position (Figure 22), was inserted upstream of MS2-RNA by adding the guide sequence with the forward primer of PSL-MS2-6X (pCS2+guide-MS2-RNA) and was ligated with the pCS2+Only vector plasmid under the control of the CMV IE-94 promoter. MS2 6X means the repetition of the sequence of MS2 stem loop part for six times. There are also 12X and 24X MS2 stem loop available, but for this research I have used the 6X MS2 stem loop.

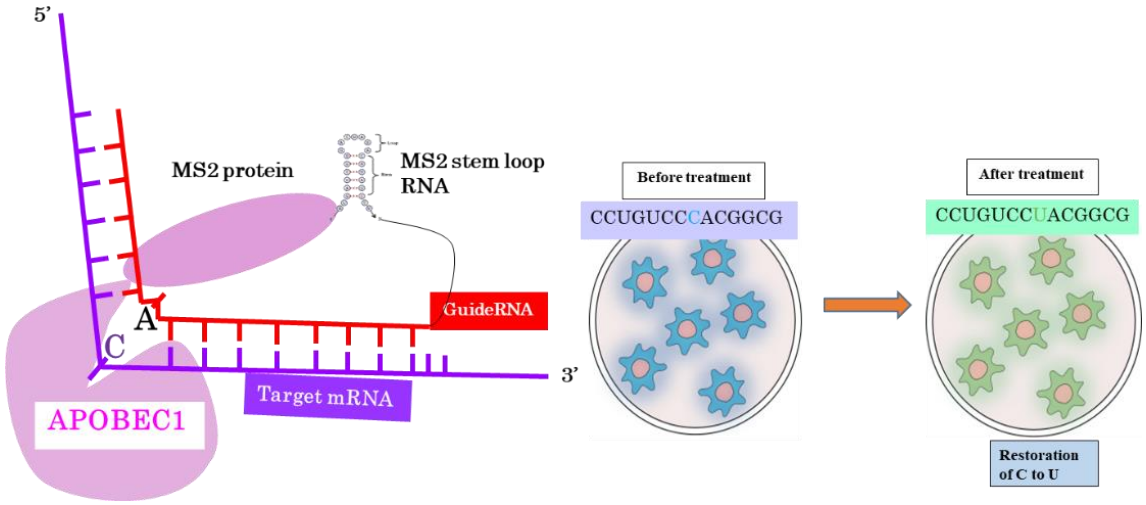


Fig 22. Schematic diagram of the MS2 system and the outcome of the experimental outline

The sequence used for this purpose was as follows: atca**GAATTC***CACTGCACGCCGTTGGACAGG*GAATGGCCATG. The atca tetra-nucleotide is a leader sequence allowing proper recognition by the restriction enzyme; bold font indicates the restriction site; italic font represents the 21 ntd guide; and the underlined portion indicates the forward primer for MS2-6X. Positive colonies were picked and confirmed by Sanger sequencing. The gRNA was chosen based on previous works done by our group with ADAR1.

Previously, I used ADARs to restore the wild-type sequence at various mutant stop codons. In that work, I designed guides of different lengths: 19, 21, or 23 ntd (upstream and downstream). I found that a 21 ntd guide was more efficient than 19 or 23 ntd versions, and that the 23 ntd guide created some off-target effects that were not detectable when the shorter sequences were used. In light of its optimal efficiency and minimal off-target effects, I used the 21 ntd gRNA in this study (2, 4). I have used this knowledge of the previous experiment for designing the guideRNA as I have used the same MS2 system in this case as well but the deaminase has been changed because of the change of the mutation type. Here, I have used APOBEC1 deaminase instead of ADAR1 as I have chosen T-to-C mutation instead of G to A mutation. Regarding the location of the target whether it should be in the 5' (as I have done) or middle or 3' position of the guideRNA, I have tried with different positions and I have found that the best efficiency I got from the 5' position of the target in the guideRNA.

The guideRNA expression was driven by the RNA Pol II promoter (CMV-IE94), which is also expressed in the human cells.

3.2.4 Culturing cells and transfection:

Stably BFP-expressing HEK 293 cells (50–70% confluent) were transfected by using Lipofectamine 2000 (Invitrogen, Carlsbad, CA, USA); 3×10^5 cells were seeded in each well of a 12-well plate. Each well contained 1 mL Opti-MEM (Gibco) and 4 µL of Lipofectamine 2000, and was transfected with 800 ng of APOBEC1 deaminase and 700 ng of gRNA into each well. The samples were then cultured in an incubator at 37°C for 6 hours. After 6 hours, the Opti-MEM was used replacing D-MEM (Gibco) to optimize growth of cells, and the cells were incubated at 37°C for 48 hours prior to observation.

3.2.5 Observation of fluorescence by confocal microscopy:

Cells were observed on an FV1000D confocal laser-scanning microscope (Olympus, Shinjuku-ku, Tokyo, Japan) under optimized conditions. To obtain very clear images, our conditions were designed to increase the effective resolution, dye selection, determination of the exposure time as well as the adjusted magnification, which enabled us to determine the exact location of the fluorescence in our cell samples, and particularly within the single cells. The intensity of fluorescence was measured by using the Fluo Viewer 10 (FluoView 10) ASW 4.0 software (Olympus). The fluorescence intensity percentage was calculated for restore GFP (conversion from BFP to GFP) and wild type GFP.

3.2.6 RNA extraction and cDNA synthesis from transfected cells:

Following microscopic observation, cells were harvested from the dish, and extracted total RNA by using the TRIzol reagent (Invitrogen). cDNA was synthesized by using the SuperScript III First-Strand Synthesis System (Invitrogen) from the extracted RNA for RT-PCR.

3.2.7 Confirmation of sequence restoration by PCR-RFLP:

To confirm successful restoration of the wild-type sequence, PCR products amplified using GoTaq polymerase (Promega, Madison, WI, USA) on a GeneAmp PCR system 9700 (Applied Biosystems, Foster City, CA, USA) were digested by using a restriction enzyme that distinguished between the edited and unedited DNA sequences. The amplicons were subjected to run in 6% polyacrylamide gel electrophoresis followed by staining with SYBR Green dye (Invitrogen). 100 ng of cDNA was used for each PCR reaction, the total reaction volume was 20 ul. After the PCR amplification 10 ul of PCR product was taken for restriction digestion, where incubation was done at 37°C for 2

hours with *Btg*I (New England BioLabs) restriction enzyme, which cleaved the BFP sequence into two fragments of 201 and 123 bp but left the restored GFP sequence undigested. During the gel electrophoresis equal volume (2 ul) of digested product was loaded into the 14 well comb. Imaging was done using the LAS 3000 imager.

The presence of the intact 324 bp sequence confirmed restoration of C to U in the mRNA (i.e., conversion of BFP to GFP). Band intensity was measured using the Image J software (NIH, https://imagej.net/Citing) and editing efficiency was calculated from the band intensities using the following equation:

$$\frac{\text{Uncut T at 324 bp}}{\text{Uncut T at 324 bp + Cut C at 201 bp + Cut C at 123 bp}} \times 100\%$$

3.2.8 Sanger's sequencing:

After doing PCR with equal amount of cDNA (100 ng) in each reaction of 20 ul volume, the PCR products were resolved on 1% agarose gels, the bands were cut out and frozen. DNA was purified using the QIAquick Gel Extraction kit, and concentration was measured on an ND-1000 spectrophotometer. Sequencing of the purified DNA was performed using the Big Dye Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Technologies, Waltham, MA, USA) using the forward and reverse primers for GFP. The raw sequencing data were analyzed using the Sequence Scanner software, version 2 (Applied Biosystems). 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 previous works on calculation of editing efficiency from peak area and peak height, I also calculated by measuring the area (20) and peak height (37, 38) using the ImageJ

software (NIH, https://imagej.net/Citing).

Editing efficiency (sense) =

$$\text{Considering Area: } \frac{\text{Area of T}}{\text{Area of T + Area of C}} \times 100$$

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

The chromatogram height has also been measured and thus the editing rate has been analyzed using the Edit-R software according to the instruction given by the Kluesener group (49).

3.2.9 Total RNA-sequencing (RNA-seq):

RNA-seq analysis was performed by Filgen (Nagoya, Japan). Read trimming to eliminate low-quality reads from the analysis was performed using the CLC Genomics Server 11 (Qiagen). Reads with the following properties were removed: bases with Phred score < 13; more than three 'N' nucleotides; length below 100 bp. Mapping was performed using the RNA-seq analysis function.

RNA-seq mapping parameters were as follows:

Reference type = Genome annotated with genes and transcripts

Reference sequence = *Homo sapiens* (hg38) sequence + BFP1 sequence

Gene track = *Homo sapiens* (hg38) (Gene) + BFP1 gene

mRNA track = *Homo sapiens* (hg38) (mRNA) + BFP1 mRNA

Mismatch cost = 2

Insertion cost = 3

Deletion cost = 3

Length fraction = 0.8

Similarity fraction = 0.8

Global alignment = No

Auto-detect paired distances = No

Strand specific = Both

Maximum number of hits for a read = 10

The average coverage of the coding region in the mapping results was 82.6 for HeK_293T3F (edited GFP sample after application of two factors in BFP stable cells) and 7.5 for BFP_1 (targeted mRNA of BFP stable HEK 293 cells), and 85% of sequence reads were mapped onto exons.

Basic Variant Detection in CLC Genomics Server 11 was used for variant detection:

1. Minimum coverage = 10 (lower limit for the number of reads mapped to the mutation location).

Minimum count = 5 (lower limit for the number of reads actually calling the mutation)

Minimum frequency (%) = 1.0 (lower limit for the percentage of reads that mapped with mutations, calculated as counts / coverage).

Mutations that met all three conditions were detected.

For analyzing the RNA seq result Jitter plot was used for detecting the off target events following the process of Grunewald et al., group [23].

Box plotting was also done by using the RNA seq analysis data, which showed the median of the Ratios of edited and unedited, it means the target BFP (mutated GFP) and restored GFP.

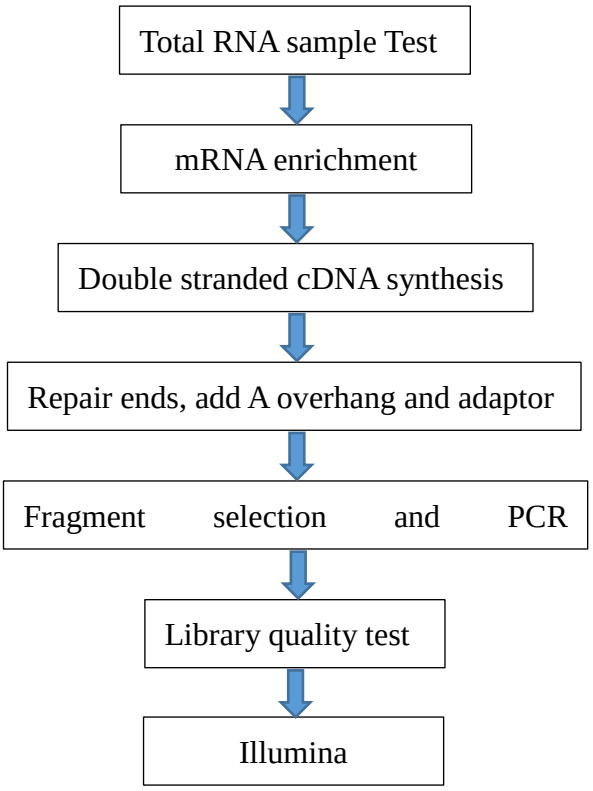


Fig 23. Flowchart showing the workflow of the RNA-seq analysis. From the RNA samples to the final data, each step, including sample test, library preparation, and sequencing, influences the quality of the data, and data quality directly.

3.3 RESULTS

3.3.1 Restoration of RNA using an artificial enzyme system:

For the purpose of confirming the editing activity of our developed artificial enzymatic RNA editing system, I used human embryonic kidney (HEK) 293 cells stably expressing BFP, a point-mutant derivative of GFP. In these cells, restoration of the original GFP sequence could be readily visualized as a change from blue to green fluorescence. To actually restore the sequence, I needed two more factors: the gRNA and the APOBEC1 deaminase, both the constructs were prepared under the control of the pol II CMV IE-94 promoter in the pCS2-MT and pCS2+ only expressing vectors, respectively.

The bacteriophage has the characteristics that its coat protein can tightly bind with the stem loop and this tight interaction aids in specific recognition of the nucleotides within the mRNA. I took the advantage of this phenomenon, regarding tight junction between the coat protein and stem loop, by combining the MS2 coat protein with the APOBEC 1 and MS2 stem loop with the gRNA, which in turn guided the deaminase to reach the specific target nucleotide and the whole system showed the activity by editing C-to-U.

3.3.2 LSM confocal microscopy:

For the confirmation of the editing of our developed artificial enzymatic system, I observed the expression of the fluorescence in the BFP expressed HEK 293 cells which I observed via fluorescence microscopy. To obtain more clear images, I observed by using an LSM confocal microscope (Figure 24b). When only one factor was transfected (e.g., either only APOBEC1 or only gRNA) into the BFP stably transformed cells, only blue fluorescence was observed but not green fluorescence, which proved that no restoration had taken place (Figure 24a). When both the factors were transfected

together (APOBEC1 + gRNA) into the same BFP stably transformed HEK 293 cells, green fluorescence was detected. More precisely, expression of BFP fluorescence to GFP was restored in cells, where the genetic code was restored because of the editing of C-to-U, and most of the fluorescence was in the cytosols (Figure 24b).

Percentage of restoration was calculated based on the fluorescence intensity of the restored GFP where I considered BFP intensity as 100%. As I have done the transfection into the BFP stably transformed HEK 293 cells so I considered the BFP fluorescence expression as 100%. Regarding the process I obtained 39% editing percentage. It means 39% of the BFP was restored to GFP by C to U RNA editing (Figure 24c).

Moreover as a positive control I selected wild type GFP because after the restoration of the genetic code (C-to-U) BFP will convert to GFP and also for validating the transfection efficiency I used wild type of GFP. For the above mentioned purposes, I transfected the wild type GFP into the BFP stably HEK 293 cells. Along with the restored GFP I also measured the GFP fluorescence intensity in wild type GFP case as well calculated the transfection efficiency by considering the BFP as 100% intense. Wild type GFP intensity was 72% (Figure 24c).

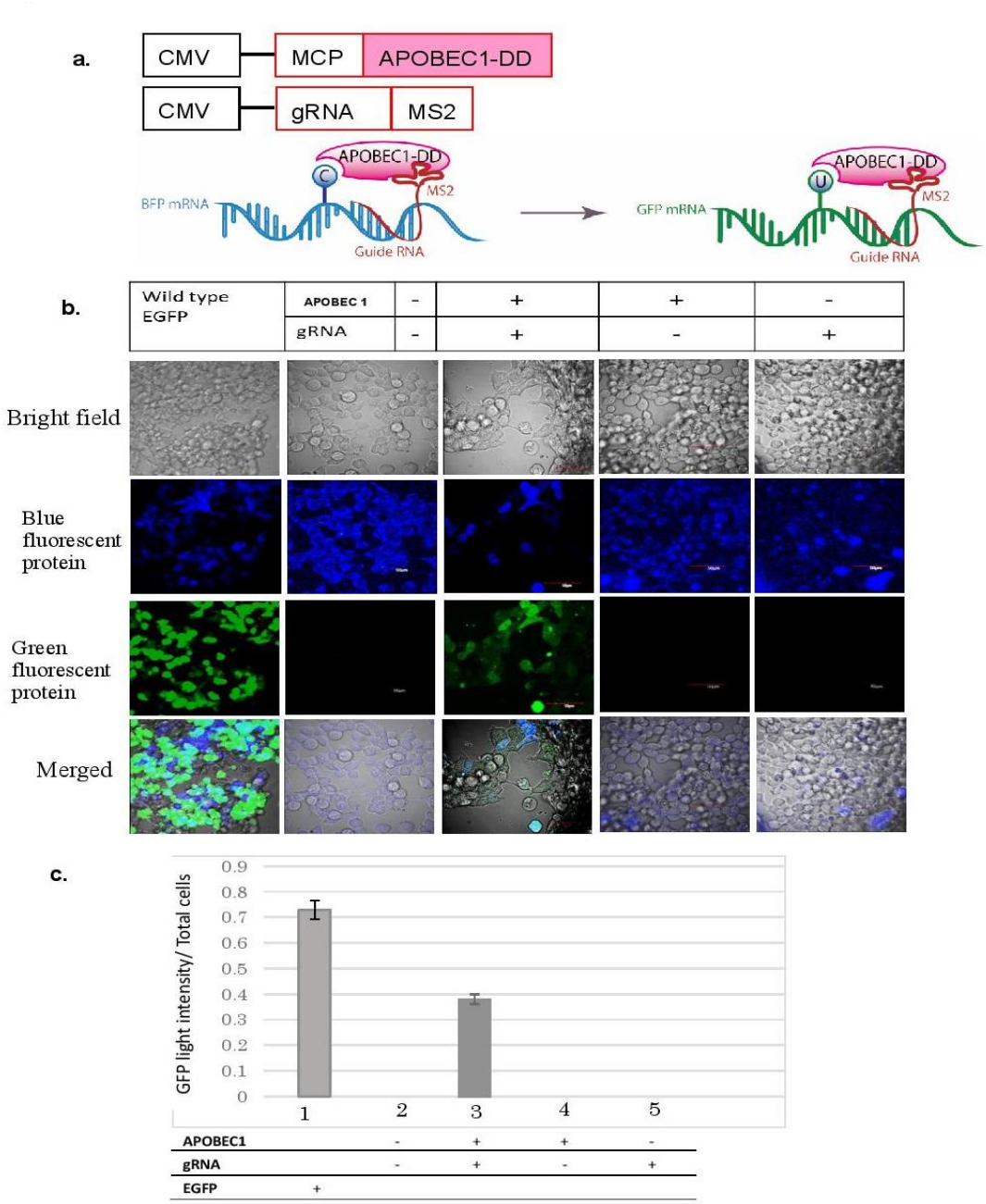


Fig 24. a. Schematic diagram of the MS2 system, The MS2 coat protein is attached to the APOBEC 1 catalytic domain and the MS2 stem loop is attached to the gRNA under the control of the CMV promoter. The MS2 coat protein and stem loop can bind to each other, enabling detection of a specific nucleotide sequence within the mRNA b. Stably BFP expressing HEK 293 cells were transfected with wild type of GFP, one factor (e.g., either APOBEC1or only guide RNA) or two factors (APOBEC1 + guide RNA). Green fluorescence expression was observed only when two factors were present, implying that both factors were necessary for C-to-U editing. Imaging was performed by **LSM confocal microscopy**. C. Calculation of the fluorescence intensity by using the Fluoviewer 10 software and for all data the statistical analysis (mean±SEM) was calculated, (n=5)

3.3.3 Confirmation of the restoration by PCR-RFLP

Target specificity is an important consideration in the context of genetic code restoration. Therefore, to verify specificity at the sequence level, I conducted reverse-transcriptase PCR followed by restriction fragment length polymorphism (RFLP) analysis of the BFP and GFP genes using the *Btg*I restriction enzyme. *Btg*I can only cut BFP, but neither the wild type GFP nor restored GFP (Figure 25a). The total length of the PCR amplified fragment was 324 bp; digestion with *Btg*I yielded two bands at 201bp and 123bp. According to the sequence preference of the enzyme, only the BFP was cut, whereas the wild type GFP and restored GFP remained undigested at 324bp. When GFP was restored, the two shorter bands were still present at 201bp and 123bp because the whole transfection process was done into the BFP stably transformed HEK 293 cells (Fig 25b). And, 100% of editing is near about impossible as a result there were still remaining part of BFP.

When only one factor (either APOBEC1 or gRNA) was transfected into the BFP-expressing HEK 293 cells, the PCR fragment was completely digested by the restriction enzyme, and there was no remaining band at 324bp. Consistent with the confocal microscopy observations, neither of the two factors on its own could restore BFP to GFP; consequently, almost 100% of the PCR fragment of BFP was cut by the *BtgI* enzyme.

Band intensity was not the same for all the samples. For the uncut bands of BFP only BFP, both the bands at 123 and 201 bp were strong but when BFP was transfected with only one factor either APOBEC1 or guideRNA the band intensity decreased. Although in case of the restored sample where both the editing factors were transfected together, remained uncut band was found at 324 bp (Fig 25b), if the intensity is compared with the two other bands at 123bp and 201bp of the same lane of sample, I could see overall

the intensity was decreased.

In case of BFP alone (without any editing factor such as APOBEC 1 or gRNA), the fragment was completely digested by the restriction enzyme. The band at 324 bp was only present when the genetic code was restored to GFP (conversion from BFP to GFP, C-to-U), i.e., when the cells were co-transfected with gRNA and APOBEC1, only then the genetic code was restored from BFP to GFP, but still two small bands were there at 201bp and 123 bp, as within a single cell 100% editing is not possible and our data also proves that.

I did the densitometry analysis of the band intensity from the PCR-RFLP PAGE electrophoresis bands. Calculation of the PCR-RFLP band intensity confirmed that BFP had been restored to GFP. To standardize the data, the PCR-RFLP experiment was repeated several times (N=5), the intensity was measured by ImageJ software, and editing efficiency was calculated individually in each replicate. Ultimately, the mean editing efficiency was 20.4% (0.204±0.05), (Fig 25c**).**

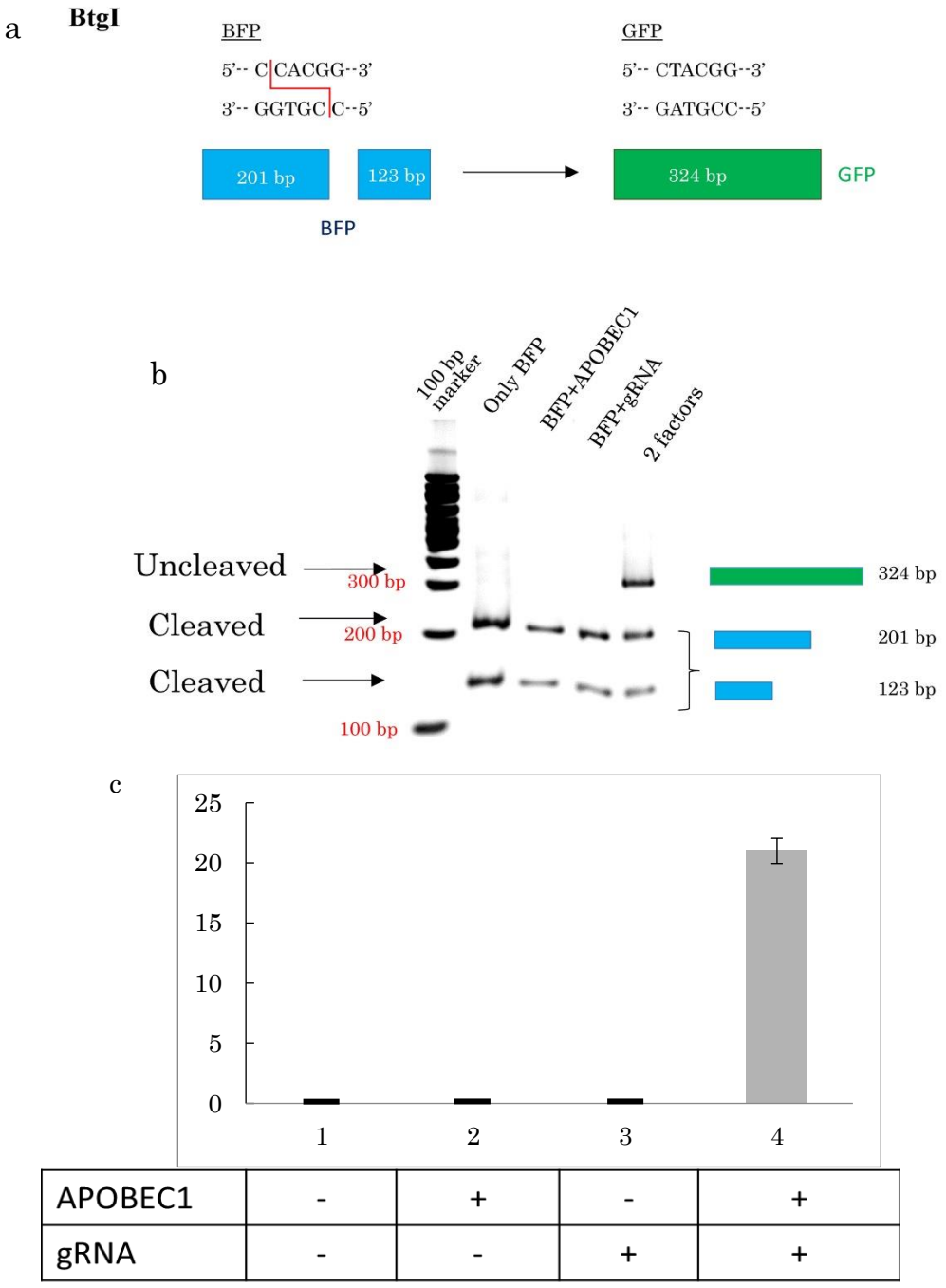


Fig 25. a. Schematic illustration of RFLP. The *Btg*I restriction enzyme can cut the BFP but not wild type of GFP or restored GFP (C to U converted) b. PCR-RFLP of cDNA extracted from transfected cells (HEK 293 stably expressing BFP), restriction-digested with *Btg*I. BFP was cleaved into two fragments, 201 and 123 bp, whereas restored GFP was not cleaved and remained at 324 bp. c. For all data the statistical analysis (mean±SEM) was calculated, where n=5.

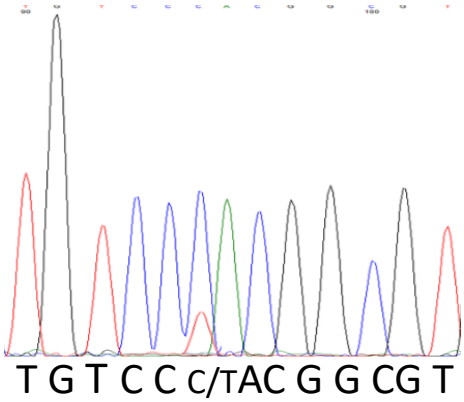
3.3.4 Confirmation by Sanger sequencing:

For more evidence of restoration the sample I also performed the Sanger's sequencing, which I did by using the sense primers. Dual peaks were only observed when cells were transfected with both the editing factors (edited T and unedited C: sense) (Figure 26 a). In case of both the sense primer, sequencing of PCR-amplified cDNA extracted from the transfected cells, confirmed that editing was successfully done and no off-target editing had occurred in the surrounding cytosines, 21 bp upstream and downstream of the target.

Editing efficiencies, calculated based on peak area and peak height were 21.5% and 21%, respectively, consistent with the value obtained by PCR-RFLP result. This experiment and the whole analysis were performed five times (N=5) under identical conditions, and similar results were obtained each time.

The peak chromatogram has also been analyzed by Edit R software through which I have also obtained 21% editing which is the identical result that I calculated via the peak height and peak area calculation from the Sanger's sequencing peak analysis (26 b).

a. Forward primer (Sense primer)



b.

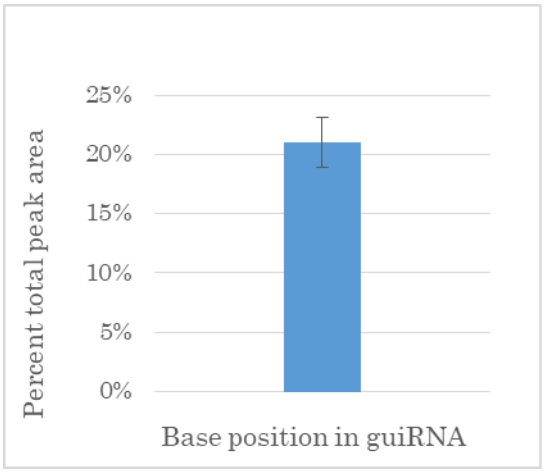


Fig 26. Confirmation of the restoration of BFP to GFP (C to U) by Sanger's sequencing. a. Forward or Sense primer (CCA to CTA). In case of sense primer the dual peaks were observed, which were due to the restoration of the genetic code from the C to U (BFP to GFP), after the application of the two editing factors (APOBEC1 deaminase and gRNA). c. Edit-R analysis of the Sense and antisense electropherogram height of the edited part of the peak and the statistical analysis (mean±SEM) has been done, where the n=5. Edit R analysis has been done by using the Sanger's sequencing Ab.1 file

3.3.5 Confirmation of sequence restoration and observation of off-target effects by total RNA-sequencing (RNA-seq):

Next, I performed whole RNA sequencing (RNA-seq) to evaluate off-target effects. The total number of reads before trimming were 30,554,972 for BFP_1 HEK 293 (Control: BFP-expressing HEK 293, targeted mRNA) and 28,523,210 for HEK_293 (Tested: in which GFP was restored after transfection with both APOBEC1 and gRNA); 30,482,237 and 28,459,029 reads, respectively, remained after trimming (Supplementary Data S2 and S3). The total number of reads sequenced did not differ between cell lines, so I can assume that any difference in the amount of expressed mRNA over the total length of the BFP gene was correlated with the number of mapping reads.

Each sequence was compared with the human genome database to identify gene

nucleotide substitution. Since human genome sequence did not including GFP gene, so I actually used the human data base+GFP.

In the BFP_1 (Control) mapped reads, a mutation with two or fewer reads that support a mutation, was identical to the mutation detected in tested HEK 293 cells (restored GFP). All the detected mutations that were found in both the control and tested samples were presented in Figure 27a also represented the mutation reads as a jitter plot, where a single dot represented a single mutation at different positions and the Y-axis represents the RNA editing percentage of those mutations at that particular position. From the analysis I could find that off target effect is almost negligible. Because the dot plotting of the experimental sample showed that the forward (C-to-U) mutation numbers are less than other mutation numbers and the pattern is also same. As a result according to the genome variance there is no off target effect. Even if I compare the numbers among the forward and reverse mutation patterns I could find that even then those are almost identical. The plotting has been done by using the analysed value of the RNA seq result and here the BFP HEK 293 cell the target sample was used as a control which was subtracted from the restored GFP in HEK 293 cells. Box plotting was also done where the median of the edited and unedited ratios was close to 1. That means off target event is almost null (27b).

In HEK_293 derived reads, the number of reads that mapped, among them very small amount of reads were not present in the BFP_1 (control). Among the mutations detected in HEK 293 (tested, restored GFP), 9053 was not present in BFP_1, including 396 C-to-T mutations and one CC-to-TT mutation..

On the other hand, 5910 of the reads obtained from BFP_1 (target mRNA, control) cells mapped to BFP. For 199C, in HEK_293 (restored GFP, tested) derived reads, a C-to-T change occurred in one of the two reads mapped to base 199. On the

other hand, in reads derived from BFP_1 (control) cells, I detected one C>A substitution, one C>G substitution, and 1608 un-mutated C nucleotides.

In addition, in the BFP_1 (control), I observed T>C mutations at 199th (Targeted site) position which was not found in case of the HEK_293 derived reads as this sample was the restored one, BFP to GFP. As a precaution, I removed the human genome sequence and mapped the reads of each cell line to the BFP gene sequence only, but the situation did not change.

Further, I looked for off-target effects. A total of 18567 mutations were present in BFP_1, whereas only 9053 mutations were present in HEK_293, suggesting that off-target effects are not a major problem in the developed artificial enzyme system. The off target effects are not significant (Figure 27c).

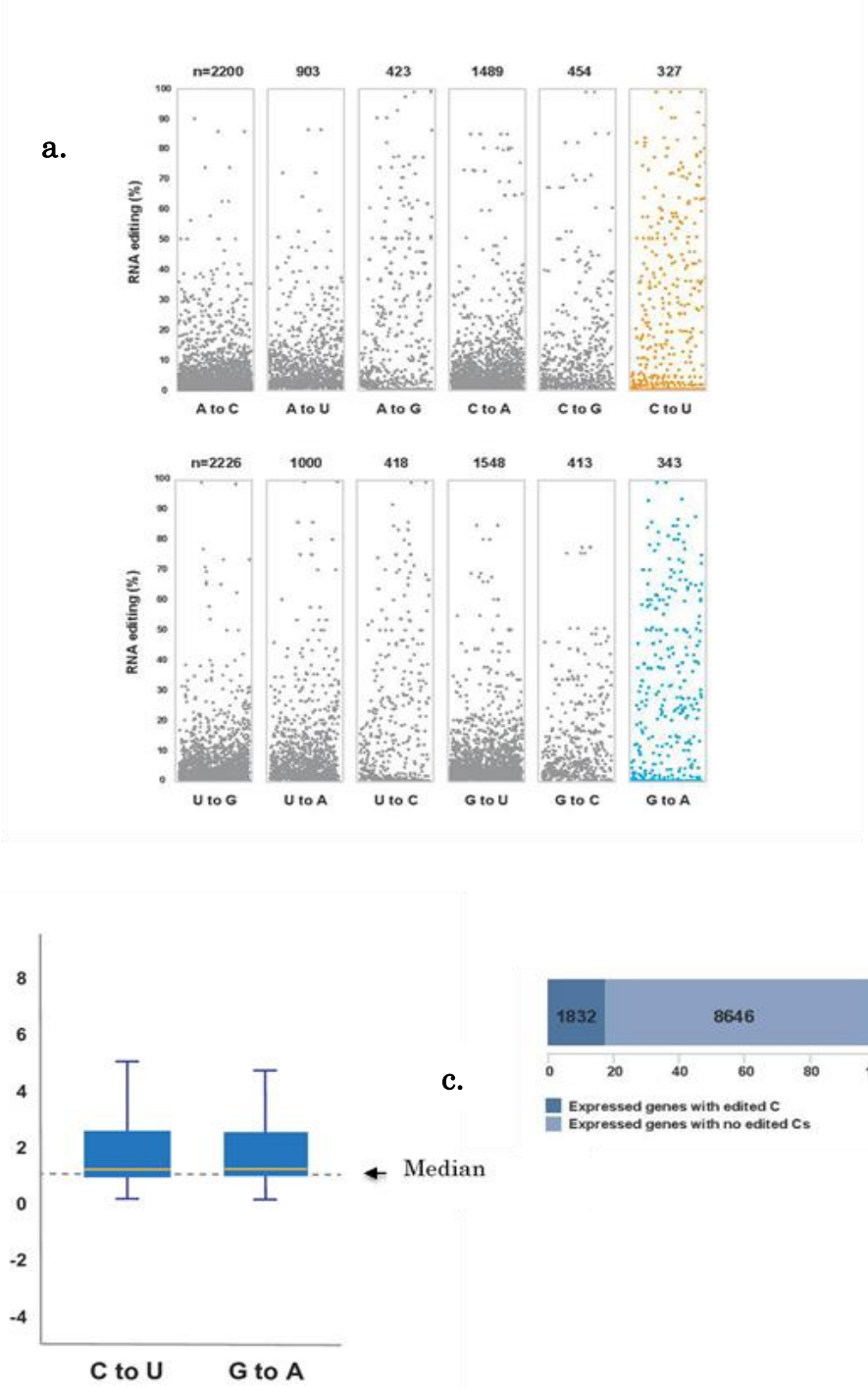
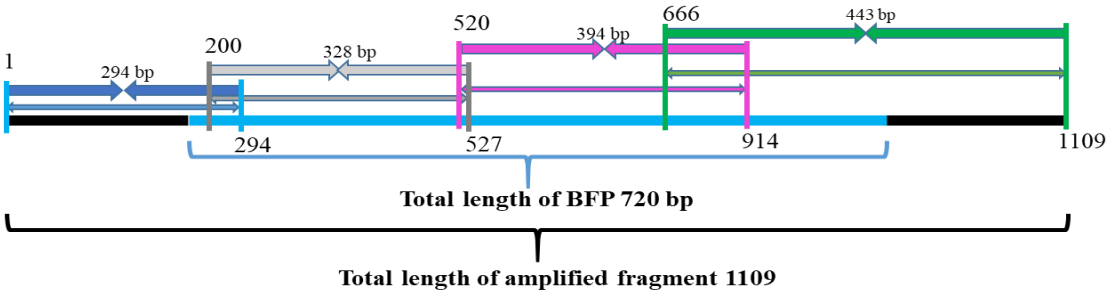


Fig 27. APOBEC 1(DD)-MS2 system induces some off-target C-to-U RNA editing in HEK293 cells. a. Percentages of expressed genes with at least one edited cytosine (C-to-U or G-to-A) in total SNVs. b. Box plots showing rate of cytosines edited by APOBEC1(DD)-MS2 compared to control (mutated BFP target in HEK 293 cells), yellow line is median. c. Jitter plots showing transcriptome-wide efficiencies of C-to-U or G-to-A edits (y-axis) identified from RNA-seq experiments in HEK293 cells modified by APOBEC1 (DD)-MS2 vs editing- negative control (BFP target, stably transformed in HEK 293 cells) . n: total number of modified cytosines identified

To confirm off target event appearing, I later on did the whole BFP sequencing by placing primers at different positions and there was also overlapping positions. The total length of the BFP was 720 bp and I amplified 1109 bp by keeping the targeted 720 bp at the center. Then Sanger's sequencing of the each PCR product was done and where it was found that only at the position 200-527 bp there was an off-target event. Which was located at the upstream of the targeted C to be edited. But it's a silent mutation, so the amino acid is not changing due to the off target event.



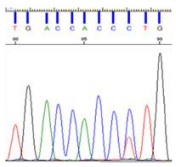
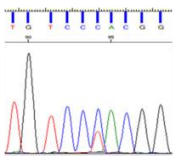
Oligos	Position range	Total amplified length	Position of off-target	Target editing
Oligo 1 Fw: GCTTATCGAAATTAATACG Rev:CACGGGCAGCTTGCCGGTGGT	1-294	294 base pair	No off target effect	There was no targeted position to be edited
Oligo 2 Fw: GCCACAAGTTCAGCGTGTC Rev: TCCTCCTTGAAGTCGATGC	200-527	328 base pair	Only the difference between the target and off target nucleotide is 5 Silent-mutation 	
Oligo 3 Fw: ACAACAGCCACAACGTCTATA Rev:AAGGCACAGTCGAGGCTGATC	520-914	394 base pair	No off target effect	There was no targeted position to be edited
Oligo 4 Fw: GACCACTACCAGCAGAACACC Rev: CAGCATGCCTGCTATTGTCTT	666-1109	443 base pair	No off target effect	There was no targeted position to be edited

Fig 28. Whole BFP sequencing by placing different primers at different positions and also there were overlapped positions. In total 1109bp were amplified, among these at the position of 200-527 only at this position one off target event was found, which was located upstream of the targeted C which was to be edited

DISCUSSION

In this study, I successfully established an artificial deaminase system based on APOBEC1 and showed that it functions as designed in C-to-U RNA editing. I also showed that the MS2 system is compatible with the design of an RNA editing enzyme model. The experimental data suggest that the deaminase domain of APOBEC1, which lacks the double-stranded RNA-binding domain (dsRBD), can be made competent for RNA editing by addition of an artificial gRNA. Because the deaminase domain of APOBEC1 lacks both a nuclear localization signal and a nuclear export signal, the engineered enzyme system should be localized to the cytoplasm, where it is likely to encounter the target mRNA. However, it remains to be determined whether small amounts of the MS2-fused deaminase chimera also localize to the nucleus.

Montiel-Gonzalez *et al.,* (2016) (30) used the Lambda N system for site-directed RNA editing along with the ADAR adenosine deaminase. In this study, I used the MS2 system, which is more efficient and has been used more frequently than the Lambda N system in RNA studies; in respect of the editing efficiency MS2 is stronger than Lambda-N system and efficient. The binding affinity towards the hairpin is 10^{-8} M for the Lambda N system, whereas the affinity towards the AUUA stem loop sequence is 10^{-9} M for MS2, which improves to 10^{-10} M when the AUCA stem loop sequence is present. This explains why MS2 is stronger than the Lambda N system (40).

Concerning the restoration of genetic code with the developed artificial deaminase enzyme system only when BFP-expressing cells were transfected with two factors, the catalytic domain and the gRNA, I could detect conversion of (mutant) BFP to GFP. No GFP fluorescence was observed when one of the factors either APOBEC1 or gRNA was transfected alone, indicating that the other cellular factors are not involved in the editing.

Further optimization of the system, including the gRNA sequence, the usability freedom of the MS2 moiety with the enzyme, and the relative proportions of the two factors, should be performed in future work. Appropriate concentrations of enzyme and reporter substrate can eliminate minor auto-editing (31, 32).

The PCR-RFLP analysis confirmed that after the application of both the deaminase and gRNA, the restored GFP was not cleaved by *Btg*I, whereas the BFP stably transformed cells which were transfected with either APOBEC1 or gRNA, these were cleaved into two fragments of 201 bp and 123 bp. This is consistent with the sequence preference of the restriction enzyme *Btg*I, and completely supports the results of Luyen *et al.*, (2015, 2017) (33, 34). These observations prove that APOBEC1 and gRNA together can restore the GFP sequence by RNA editing from the BFP transcript.

Non-restored BFP was completely cleaved into the 201 and 123 bp fragments, and no 324 bp fragment remained. By contrast, the restored GFP was not cleaved by the enzyme. A remained 324 bp band was observed due to the editing where both the enzyme and gRNA were applied together, however, still I could see the cleaved bands at 201 bp and 123 bp in the same sample of same lane where restored GFP was expressed (Figure 25b), because 100% restoration of the genetic code is not possible which I have also observed from our confocal images as well. I calculated the editing rate by densitometry analysis from PAGE band intensities, which I analyzed using the ImageJ software (NIH), and found that the mean editing efficiency was 20.4%.

Further confirmation of the developed system's ability was obtained by Sanger's sequencing. The sequencing analysis revealed that the restored GFP gave a peak corresponding to wild-type GFP: TCCTACGG instead of TCCCACGG with the sense primer. The dual peak could be observed only when the restoration of the genetic code happened due to editing. I calculated the editing efficiency based on peak height for

both the sense and antisense primers. This calculation indicated that 21.5% of BFP was edited using our deaminase approach.

I also used the Edit-R software to analyze the sequence result. I measured the peak chromatogram and calculated the editing percentage of the particular nucleotide. The analysis showed that 21% of the C has been restored to U which is identical with the measurement that I have done by peak height and peak area. I analyzed the sense sequence data and in both the cases have got the same 21% editing rate.

BFP contains a single mutation, so I detected only a single peak at that specific site. I found no off-target editing in the 21 bp upstream or downstream of the mutated site. I believe that the frequency of off-target editing is almost zero as the read through of the targeted BFP and restored GFP (after transfection with the two factors) was very similar. Interestingly, our total RNA sequencing (RNA-seq) analysis revealed that, in HEK cells, expression of both GFP expressing mRNA and BFP protein were potentially reduced. It is possible that the gRNA may have mediated some sort of RNA interference (RNAi) activity, although, I have no corroborating data to this effect. The underlying assumptions should be validated in future studies. Although when an expression vector is introduced into the cell, an endogenous transcription factor binds to the vector promoter. Therefore, transcription of endogenous genes is totally suppressed. Necessary transcription factors are deprived.

In terms of editing efficiency, several factors should be considered; in the future, improvement of these factors could increase the efficiency. For example, I used the MS2-6X stem loop, although previous experiments suggested that the MS2-12X stem loop is much more active. Hence, a gRNA constructed using MS2-12X could increase editing efficiency. Even the 1X MS2 on the either side of the guideRNA may also have a positive impact on the editing efficiency, which has already been observed by the

Katreakar et al., but for A to I editing with the ADAR1 enzyme and CRISPR approach (39).

Another approach might be optimization of the length of the gRNA, which also determines the degree of off-target editing. In previous experiments, I found that the best length of a guide sequence for site-specific editing is 21 ntd: if I increase the length of a guide, efficiency may increase, but the level of off-target editing may also increase simultaneously. As the developed system is same with the previous one, only the difference is editing enzyme and the target mRNA that is why I have used that knowledge for designing the guideRNA for this study as well.

Transfection with two factors has a greater effect on cellular conditions than transfection with a single factor. Hence, if I can make a single construct in which the deaminase and the gRNA are encoded on the same construct, this should make transfection easier and likely also increase efficiency.

The developers of the CRISPR-Cas9–AID approach to sequence restoration reported editing efficiencies of 26.2% for rice and 53.8% for tomato (25, 36), though it was a genome editing technique. This is significantly higher than the value I achieved in the present study (~21%), which was based on combining MS2 and APOBEC1 to edit the mutated GFP/BFP mRNA *in vitro*. Notably, however, this is the first time that the MS2–APOBEC1 combination has been used for C-to-U editing to correct T-to-C mutations.

In our system, off-target effects were negligible whether the system was completely based on the RNA editing. The developers of Cas9-AID also reported a very low rate of off-target editing, 0.14–0.38%, implying that off-target effects are very rare with systems of this kind, although their system was of genome editing whereas our system is for RNA editing (25). Even the group who worked with the Adenosine base editors along with the CRISPR system they have also found low rate of indels which is only

0.1% (48), but their system convert targeted A•T base pairs efficiently to G•C approximately with 50% efficiency in human cellular system, although their system was of genome editing and applied with CRISPR-ABEs. Afterwards the Zuo group has done the Cytosine base editing *in vivo* and they have also found only zero to four indels in embryos and none of them overlapped with the predicted off-target sites (44). Here they used the base editors/cas9 and the GOTI (genome-wide off-target analysis by two-cell embryo injection) technique for detecting the off target events *in vivo*.

Moreover, to prepare the CRISPR-Cas9–AID restoration tool, Zenpei *et al.*, (2017) (25) used pmCDA1, which has a backbone of ~10 kb, while the length of the Cas9 gene is 3156 bp. Thus, in total their construct is ~13 kb in length, which may be too large to deliver in a therapeutic context. Both MS2 and APOBEC1 are smaller than the components of CRISPR-Cas9, and as vectors I used pCS2-only and pCS2-MT, which have significantly smaller backbones, ~4.5 kb. Therefore, our system is smaller, so it is easier to manipulate and deliver, and has a better editing efficacy than the Cas9-AID combination tool, although that Cas9-AID system was used for the genome editing whereas our developed system is for RNA editing.

At the upstream of the targeted C there was a single off target effect, which was found after doing the whole BFP sequencing. As APOBEC 1 affects the shRNA as a result it has some indebt nature of having off target effect than that of ADARs as those affect the dsRNA. Moreover according to the position where the off target was found there was a mimic of CCC so it could be a possibility that due to the mimic of C there was the off target effect. For this research the upstream guideRNA has been used so I think a downstream guideRNA could solve the problem and but the restoration percentage could be reduced after application of the downstream but there could not be any off target event.

CONCLUSION

The findings indicated that the developed artificial deaminase system could successfully restore gene expression by editing point mutations (U-to-C) in the RNA of patients with various diseases caused by such mutations, thereby alleviating the symptoms of their disorders by performing C-to-U editing. The ability of MS2-APOBEC1 and gRNA to edit mutated RNA sequences could be used to overcome genetic diseases caused by T-to-C mutation. Future studies should seek to optimize editing efficiency, e.g., by using gRNAs of different lengths and altering the MS2 stem loop. If possible, the practical applications of this system should be tested in animal models.

REFERENCES

1. Heidenreich M., and Zhang, F. Applications of CRISPR-Cas systems in neuroscience. Nat Rev Neurosci., 17(1): 36–44, (2016).
2. Katrekar D., and Mali P. In vivo RNA targeting of point mutations via suppressor tRNAs and adenosine deaminases. Cold Spring Harbour (BioRxiv). DOI: http://dx.doi.org/10.1101/210278doi (2017).
3. Kim Y.G., Cha J., and Chandrasegaran S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proceedings of the National Academy of Sciences. 93(3): 1156–1160, (1996).
4. Porteus M.H., and Carroll D. Gene targeting using zinc finger nucleases. Nat Biotechnol, 23(8): 967–973, (2005).
5. Zhang F., Le C., Simona L., Sriram K., George M.C. and Paol A. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. Nat Biotechnol, 29(2): 149–153, (2011).
6. Hsu P.D., Lander E.S., and Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. Cell, 157(6): 1262–78, (2014).
7. Thomas G., Shannon J.S., Sai-lan S. and Jia L. Genome editing technologies: Principles and applications. Cold spring Harbor Perspectives in Biology, 8:a023754 DOI: 10.1101/cshperspect.a023754 (2016).
8. Gott J.M., and Emeson R.B. Functions and mechanisms of RNA editing. Annu Rev Genet, 34: 499-531, (2000).
9. Keegan L.P., Gallo A., and O'Connell, M.A. The many roles of an RNA editor. Nat Rev Genet, 2(11): 869–78, (2001).
10. Aisling M.F., and Andrew R.G. Adenosine deaminase deficiency: A review.

Orphanet: Journal of Rare diseases, 13: 65, DOI: https://doi.org/10.1186/s13023-018-0807-5 (2018).

11. Lane D.A., Kunz G., Olds R.J., and Thein S.L. Molecular genetics of antithrombin deficiency. Blood Rev., 10(2): 59–74, (1996).
12. Olds R.J., Lane D.A., Ireland H., Finazzi G., Barbui T., Abildgaard U., Girolami A., and Thein S.L. A common point mutation producing type 1A antithrombin III deficiency: AT129 CGA to TGA (Arg to Stop). Thromb. Res., 64 (5): 621–625, (1991).
13. Hirschhorn R., Tzall S., Ellenbogen A., and Orkin S.H. Identification of a point mutation resulting in a heat-labile adenosine deaminase (ADA) in two unrelated children with partial ADA deficiency. J Clin Invest, 83(2): 497–501, (1989).
14. Maas S., Rich A., and Nishikura, K. A-to-I RNA editing: Recent news and residual mysteries. J Biol Chem, 278: 1391–1394, (2003).
15. Stafforst T. and Schneider M.F. An RNA Deaminase conjugates electively repairs point mutations. Angew. Chem. Int. Ed, 51(44): 11166–11169, (2012).
16. Vogel P., Hanswollemenke A., and Stafforst T. Switching protein localization by site directed RNA editing under control of Light. ACS synthetic Biology, 6: 1642-1649, (2017).
17. 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).
18. Hiroki S., Kenichi N., Misaki S., Ryo F., Yukio M., Shoko H., Kaori S., Deepika K., Takanori Y., and Takaomi C.S. Introduction of pathogenic mutations into the mouse Psen1 gene by Base Editor and Target-AID. Nature Communications, 9: 2892, (2018).

19. 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).
20. Bhakta S., Azad M.T.A. and Tsukahara T. Genetic code restoration by artificial RNA editing of Ochre stop codon with ADAR1deaminase. Protein Engineering, Design & Selection, 31(12): 1-8, (2018).
21. Liqun L., Hongquan C., Wei X., Bei Y., Bian H., Jia W., Lijie W., Yiqiang C., Wei L., Jianying W., Lei Y., Wanjing S., Jimin G., Jiahao S., Min Z., Xingxu H., Bin S., Li Y. and Jia C. APOBEC3 induces mutations during repair of CRISPR–Cas9-generated DNA breaks. Nature Structural & Molecular Biology, 25: 45–52, (2018).
22. Blanc V., Xie Y., Kennedy S., Riordan J.D., Rubin D.C., Madison B.B, Mills J.C., Nadeau J.H. and Davidson N.O. APOBEC1 complementation factor (A1CF) and RBM47 interact in tissue-specific regulation of C to U RNA editing in mouse intestine and liver. RNA, 25(1): 70-81, (2019).
23. Grunewald J., Ronghao Z., Sara P.G., Sowmya I., Caleb A.L., Martin J.A. and Keith J. Transcriptome-wide off-target RNA editing induced by CRISPR guided DNA base editors. Nature, 569: 433-437, (2019).
24. Chester A.N.N., Weinreb V., Chrles W., Carter J.R. and Navaratnam N. Optimization of apolipoprotein B mRNA editing by APOBEC1 apoenzyme and the role of its auxiliary factor, ACF. RNA, 10: 1399–1411, (2004).
25. Zenpei S., Sachiko K., Mariko T., Rie T., Takayuki A., Hisaki I., Hiroshi T., Tsuyoshi Y., Hiroki K., Kenji M., Hiroshi E., Keiji N., Tohru A. and Akihiko, K. Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. Nature Biotechnology, 35: 441–443, (2017).

26. Mali P., Luhan Y., Kevin M.E., John A., Marc G., James E.D., Julie E.N. and George M.C. RNA-guided human genome engineering via Cas9. Science, 339(6121): 823–826, (2013).
27. 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).
28. Johansson H.E., Liljas L. and Uhlenbeck O.C. RNA recognition by the MS2 phage coat protein. Sem Virol, 8(3): 176–185, (1997).
29. Bertrand E., Chartrand P., Schaefer M., Shenoy S.M., Singer R.H. and Long R.M. Localization of ASH1 mRNA particles in living yeast. Molecular Cell, 2(4): 437–445, (1998).
30. Montiel-González M.F., Vallecillo-Viejo I.C. and Rosenthal J.J. An efficient system for selectively altering genetic information within mRNAs. Nucleic Acids Res, 44 (21): 157-168, (2016).
31. Fukuda M., Umeno H., Nose K., Nishitarumizu A., Noguchi R. and Nakagawa H. Construction of a guide-RNA for site-directed RNA mutagenesis utilizing intracellular A-to-I RNA editing. Sci. Rep, **7**: 41478:1-13, (2017).
32. Schneider M.F., Wettengel J., Hoffmann P.C. and Stafforst T. Optimal guideRNAs for re-directing deaminase activity of hADAR1 and hADAR2 in trans. Nucl. Acids Res, 42(10): e87, DOI: 10.1093/nar/gku272 (2014).
33. Luyen T.V., Thanh T.K.N., Shafiul A., Takashi S., Kenzo F., Hitoshi S. and Toshifumi T. Changing Blue Fluorescent Protein to Green Fluorescent Protein Using Chemical RNA Editing as a Novel Strategy in Genetic Restoration. Chem Biol Drug Des, 86: 1242–1252, (2015).
34. Luyen T.V. and Toshifumi T. C-to-U editing and site-directed RNA editing for the

correction of genetic mutations. BioScience Trends Advance Publication, 11(3): 243-253, (2017).

35. Luyen V.T., Ooka Y., Alam S., Suzuki H., Fujimoto K. and Tsukahara T. Chemical RNA Editing as a Possibility Novel Therapy for Genetic Disorders. International Journal of Advanced Computer Science, 2(6): 237-241, (2012).
36. Keiji N., Takayuki A., Nozomu Y., Satomi B., Mika K., Mayura T., Masao M., Aya M., Michihiro A., Kiyotaka Y.H., Zenpei S., and Akihiko K. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. Nature Communications, 353 (6305): 8729 (1-7), (2016).
37. Eggington J.M., Greene T. and Bass B.L. Predicting sites of ADAR editing in double-stranded RNA. Nature Communications, 2(319): 1-9, (2011).
38. Rinkevich F.D., Schweitzer P.A. and Scott J.G. Antisense sequencing improves the accuracy and precision of A-to-I editing measurements using the peak height ratio method. BMC Res. Notes, 5(63): 1–6, (2015).
39. Katrekar D., Genghao C., Dario M., Ashwin G., Atharv W., YuRu S., Shyni V., and Prashant M. In vivo RNA editing of point mutations via RNA-guided adenosine deaminases. Nat Methods, 16(3): 239-242, (2019).
40. Montiel-Gonzalez M.F., Quiroz J.F.D. and Rosenthal J.J.C. Current strategies for Site-Directed RNA Editing using ADARs. Science Direct Methods, 156: 16-24, (2019).
41. Komor A., Kim Y., Packer M. et al. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature, 533: 420–424. (2016).
42. Kluesner M.G., Nedveck D.A., Lahr W.S., Garbe J., Abrahante J., Webber B., and Moriarty, B.S. EditR: A method to quantify base editing via Sanger

sequencing. The CRISPR Journal, 1(3): 239-250, (2018).

43. Zhang X.H., Tee L.Y., Wang X.G., Huang Q.S. and Yang, S.H. Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. Molecular Therapy-Nucleic Acids, 4, e264. DOI:10.1038/mtna.2015.37(2015).
44. Zuo E., Sun Y., Wei W., Yuan T., Ying W., Sun H., Yuan L., Steinmetz L.M., Li Y., and Yang H. Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. Science, 364(6437): 289-292, (2019).
45. Martin A.St., Salamango D.J., Serebrenik A.A., Shaban N.M., Brown W.L., and Harris R. S. A panel of eGFP reporters for single base editing by APOBEC-Cas9 editosome complexes. Scientific Reports, 9: 497, (2019).
46. Gehrke J.M., Cervantes O., Clement M. K., Wu Y., Zeng J., Bauer D.E., Pinello L., and Joung J.K. An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. Nature Biotechnology, 36: 977–982, (2018).
47. Marx V. Base editing a CRISPR way. Nature Methods, 15: 767–770, (2018).
48. Gaudelli N.M., Komor A.C., Rees H.A., Packer M.S., Badran A.H., Bryson D.I. and Liu D.R. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. Nature, 551: 464–471, (2017)

CHAPTER IV

A study on pol II, pol III promoters and single construct (made of combination of pol II and III promoter) for restoration efficacy in case of C-to-U editing

4.1 INTRODUCTION

The term genetic engineering at first alluded to different techniques utilized for the alteration or manipulation of organisms through the processes of heredity and reproduction. As such, the term grasped both the artificial selection and all the interventions of biomedical procedures, which includes artificial insemination, *in vitro* fertilization (e.g., “test-tube” babies), cloning, and gene manipulation etc. It is the process to control the expression and functions of the intracellular target genes that have been broadly used in the fundamental research as well as medicinal and applications for treatment (1). In recent times, several other techniques for genome editing have been made possible the manipulation of the targeted genomic information (1, 2, 3, 4, 5).

There is a broad assembling of lentiviral short hairpin (shRNA) expressing vectors, which has been generated to accountable with different necessities in RNA interference technology. More specifically, continuous modification has been occurred in the promoters directing shRNA expression to ameliorate gene silencing efficiency. There are three types of RNA polymerase promoters such as type I, type II, and type III. All of them are presently in use. But among them polymerase III is the most commonly used one because it has various advantageous points such as it naturally directs the synthesis of small, highly abundant non-coding RNA transcripts, which defined termination sequences consists of 4-5 thymidines (Ts) and have no particular requirement for the downstream elements for promoting (1).

Pol III genes and promoters do not get polyadenylated and thus are ideal for making small nuclear RNAs such as shRNA hairpins. On contrary, Pol II transcription requires a poly-A site to terminate the mRNA and process it correctly, which does not fulfill the purpose of making a shRNA. This might be a reason for making the pol III such as U6

more effective than the pol II CMV promoter. Pol III promoters are popular since they allow expression of non-coding RNA molecules with precise 5' and 3' sequences (i.e. no 5' cap or 3' poly-A tail). Thus, my prediction is that the CMV promoter which recruits Pol II will not be more effective than pol III, U6.

The best example for pol III promoter is U6 promoter, usually used for shorter transcript, it synthesizes shRNAs, tRNAs, rRNA 5s etc., at an optimal level and its transcript initiation site is started with G (2-4). Short hairpin RNAs (shRNAs) transcribed by U6 promoters can trigger sequence-selective silencing of gene expression in culture and *in vivo* which may allow the analyses of gene functions or as a potential tool for therapy (5), can also avoid the transcription of irrelevant DNA sequences, thus it reduces the expected results e.g: off-target mutation, which is a major concern in case of many editing approaches (6).

U6 has a unique character, though it possesses highly efficient transcription but it is difficult to use same promoter among different distantly related species. Even differences in transcriptional activity could be considered when same U6 promoter could be used among divergent plants (7-10).

4.1.1 Single construct having CMV and U6 promoter in combination:

The discovery of RNA interference (RNAi) was the most exciting development in gene regulation, through which short interfering RNAs (siRNAs) intercede selective gene inactivation by mRNA obliteration (13).

Short hairpin RNA (shRNA) coordinated by RNA pol III or pol II promoter, demonstrated to be equipped for silencing gene expression, which allowed the investigation of gene functions or as a potential tool for gene therapy.

The shRNAs transcribed under the control of RNA pol III promoters *in vivo*, can trigger

degradation of the correlated mRNAs which is similar to siRNAs and hinder definite gene expression (18, 19). Additionally, RNA pol II plays the role for synthesis of numerous noncoding RNAs, consisting of small nuclear and nucleolar RNAs (13). Recently, siRNA transcripts expressed from a RNA pol II promoter, cytomegalovirus (CMV) promoter, have been demonstrated to be competent for reducing gene expression in mammalian cells and zebrafish (14, 15).

A pol II enhancer such as CMV promoter could be able to enhance the pol III transcription. In previous study it has been shown that the enhancer from the cytomegalovirus immediate-early promoter, when placed near the pol III U6 promoter in a plasmid vector, could enhance the activity of the U6 promoter, increase shRNA expression, and strengthen the gene-silencing effect in human cells (16).

Enhancing U6 promoter activity by CMV enhancer in target cells has been proven to be an effective way to obtain satisfactory editing efficiency (35). U6 promoter is constitutively active in a variety of cell types and maintains relatively high activity by providing approximately 4×10^5 transcripts per cell. However, under some circumstances, unmodified Pol III promoter is not sufficient to provide satisfactory depletion of target transcripts even without reported TI (34, 35, 36, 38, 39, 40). The addition of a CMV enhancer adjacent to U6 promoter or hybrid CMV-H1 promoter has been reported to improve the efficiency of RNAi (35, 36) or shRNA delivery *in vivo* (40).

Whereas other studies have shown that, attaching the CMV promoter or the enhancer at the upstream of the U6 promoter has shown an improved efficiency. There are also many previous studies who have found the highly effective activity by a combination of the CMV and U6 promoter in case of the gene specific knockdown. From these perspectives I can assume that for the case of editing efficiency the combination of the

CMV pol II and U6 pol III will provide a higher efficiency than the either only CMV or only U6 promoter (17)

It is well perceived that enhancer and promoter recognition by transcription factors, repressors, and auxiliary proteins is a complicated process, involving both essential and auxillary sequence characteristics of a gene expression regulatory element. The number, diversity, orientation and placement of transcription factor-binding sites within an enhancer and/ or a promoter are consorious parameters that define gene expression levels.

In this study, attaching the CMV promoter or the enhancer at the upstream of the U6 promoter has shown an improved effectiveness, which supports the findings of many researchers, who have found that by a combination of the CMV and U6 promoter in case of the gene specific knockdown and where the combined promoter showed the highest reduction activity comparing with the single promoter.

4.2 MATERIALS AND METHODS

4.2.1 Target plasmid (mutated EGFP or BFP) construct preparation:

The preparation of the target BFP was done by following the protocol of Luyen *et al*., 2015 (41, 42). Then the BFP construct was treated into the HEK 293 cell and maintained with G418 @ 500 ng/mL. The positive colonies were picked and confirmed by Sanger's sequencing and NCBI Blast.

4.2.2 APOBEC 1 deaminase plasmid construct preparation:

To enable the targeting of the enzymes to a BFP target codon of interest, I cloned the deaminase domain of APOBEC 1 downstream of MS2 in pCS2+MT using the XhoI and XbaI (Takara, Shiga, Japan) restriction enzymes to yield pCS2+MT-MS2HB-APOBEC 1, by PCR amplification from HEK 293 cells the forward and reverse primers consisting of particular restriction sites (XhoI catalytic APOBEC 1 Fw: tccactcgagatgccctgggagtttgacgtctt, XbaI catalytic APOBEC 1 Rv1: acggtctagattaagggtgccgactcagaaactc XbaI catalytic APOBEC 1 Rv2: acggtctagattattaagggtgccgactcagaaactc), Red color represents the restriction sites. Positive colonies were picked and confirmed by Sanger's Sequencing. By using the ExPASY Bioinformatics resource portal and NCBI-BLAST search the in frame and domain was confirmed.

4.2.3 Preparation of the guide for directing the APOBEC 1 to targeted site:

4.2.3.1 Preparation of the guide under the control of pol II CMV promoter:

For preparation of the guide RNA 21 bp nucleotides, which was complementary sequence to target mRNA site, was inserted upstream (atca**GAATTC** *CACTGCACGCCGTTGGACAGG*GAATGGCCATG) to MS2-RNA with either forward

(pCS2+guide-MS2-RNA), (here in the primer atca or attc is given for proper identification of sequence by the restriction enzyme, Bold part is the restriction enzyme site, Italic format is the 21bp guide and the underline portion is the forward and reverse primer for MS2). Positive colonies were picked and confirmed by Sanger's sequencing. The guide was randomly chosen based on some previous works.

According to previous work, I found that 21bp had better efficiency than 19 bp and 23 than 21 but in case of 23 bp guide I could find some off targets which was not found in 21 bp even not in 19 bp. So by considering the efficiency as well as the off target issue I found that 21 bp length of the guide had the best efficiency, that's why for this experiment as well I have tried to use this 21 bp guide as the best one (11, 12). Regarding the position of the target nucleotide C, a mismatch A was placed at the place of the target C and that was located at the 5th position in the 21 bp guideRNA.

4.2.3.2 Preparation of the guide under the control of pol III U6 promoter:

The guide RNA with CMV promoter was used as a vector for the purpose and U6 promoter was amplified from the pX330-U6-Chimeric_BB-CBh-hSpCas9 plasmid vector by PCR amplification. For the purpose CATATGgagggcctatttcccatgat (Fw primer with the NdeI restriction enzyme) and CacggCCTAGGggtgtttcgtcctttccaca (Reverse primer with BamHI restriction enzyme). After the amplification and restriction digestion of the U6 promoter part from the plasmid vector, it was inserted in to the guide vector with CMV promoter where the CMV was replaced with U6 by restriction digestion by NdeI and BamHI enzymes. Finally pCS2+U6+guideRNA+MS2 stem loop, this construct was prepared. The SV-40 terminator would act for U6 promoter as well.

4.2.4 Preparation of the Single construct having APOBEC 1 deaminase under the pol II CMV promoter's control and guideRNA under pol III U6 promoter's control:

For the preparation of the single construct the guide RNA construct under the control of the pol III U6 promoter was used as a vector whereas the whole portion of the APOBEC1 deaminase along with the MS2-HB coat protein part under the control of the pol II CMV promoter up to SV-40 terminator was used as an insert.

The CMV- MS2HB-APOBEC 1-SV-40 terminator portion was collected by doing the restriction digestion of the APOBEC1 construct with NdeI and HpaI restriction enzyme. After the restriction digestion both the vector and insert were made blunt by the Klenow fragment treatment by using Klenow fragment Kit (Takara). The blunt ended vector and insert were ligated together and after the ligation the direction does not matter as both the parts have their individual promoter and terminator. The final construct was pCS2+CMV+MS2HB+APOBEC1+SV40+U6+guideRNA+MS2stem loop+pCS2-Only. After transformation the positive colonies were extracted by QIAGEN Plasmid Midi Kit (QIAGEN) and the final construct was sequenced for the final confirmation and the BLAST was done

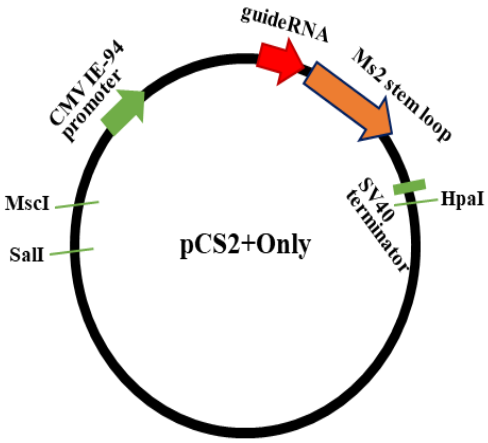


Fig 28. Guide construct under the control of the pol II CMV promoter

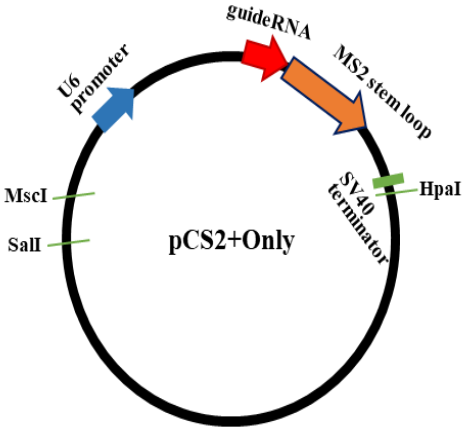


Fig 29. Guide construct under the control of the pol III U6 promoter

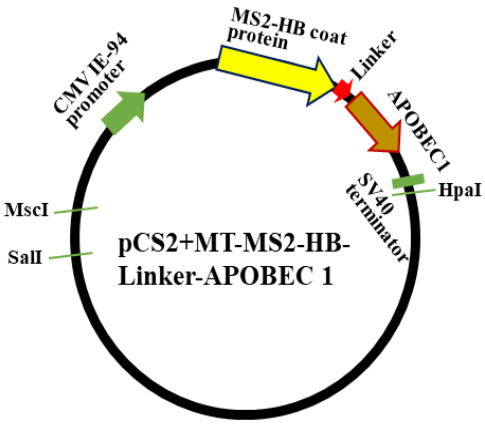


Fig 30. APOBEC 1 deaminase construct under the control of pol II CMV promoter

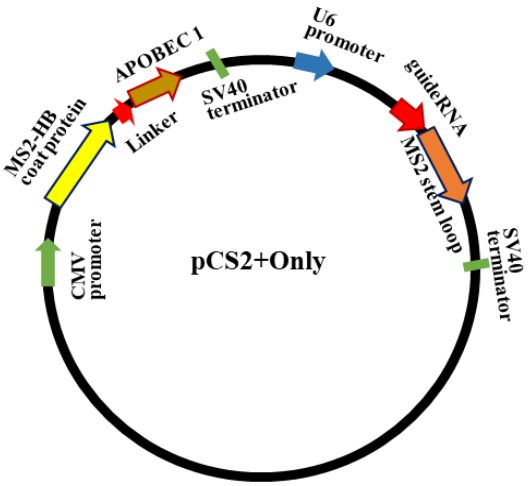


Fig 31. Single construct, the guide portion under the control of the pol III U6 promoter and the APOBEC 1 deaminase portion under the control of the pol II CMV promoter. Here CMV promoter is located at the upstream of the U6 promoter

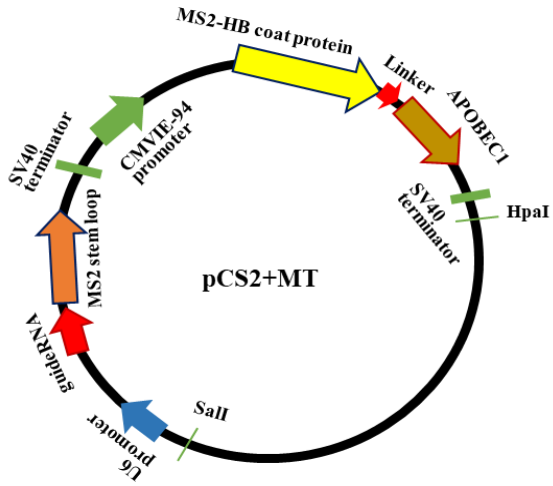


Fig 32. Single construct, the guide portion under the control of the pol III U6 promoter and the APOBEC 1 deaminase portion under the control of the pol II CMV promoter. Here CMV promoter is located at the downstream of the U6 promoter

4.2.5 Western blot analysis:

To check the expression of the APOBEC 1 in to the cells western blot was done and the interaction was also checked by the process.

Cell preparation for Transfection:

1. Preparation of 10 cm dish with 10 mL of DMEM (10% FBS), and adequate amount of cells
2. Incubating the dish for 24 hours at 37^0C and 5% CO2
3. After 24 hours check the cells and the confluency, the confluency was about 70%, so the cells were ready for transfection
4. Transfection of 15 ug of Plasmid DNA into the 1 mL of Optimem with 60 ul of Lipofectamin-2000.

   - No need to change the media (DMEM) before transfection
   - 1 mL of Optimem taken into two Eppendorf tube 500 ul each
   - 15 ug of DNA dissolved in to one of the Eppendorf tubes and let it to be in room temperature for 5 mins
   - Mix 60 ul in to another tube and let it to be in room temperature for 5 mins
   - Mix both of these together and keep it in the room temperature for 20 mins
   - Then spread it over the dish

5. After 6 hours change the media and also on the next day after 24 hours of transfection change the media once again
6. After 48 hours of transfection collect the cells for the protein extraction

   - Clean the dish with the Ice cold PBS (500 ul per dish), remove PBS
   - Give 500 ul of Lysis buffer in to the dish and keep the dish on ice for 30 mins
   - After 30 mins collect the cells in to the clean 1.5 ml tube
   - Apply sonication for 10 seconds
   - Do the centrifugation for 10 mins @ 11000Xg
   - Lysed cells will be at the bottom and supernatant having the protein

7. Mix 12 ul of supernatant taken into a new clean tube and mix with 3 ul of sample buffer (Total 15 ul)
8. Incubate the total sample in 100^0C in heat block for 3 mins
9. Then keep on ice
10. Prepare the Running gel and stacking gel (Butanol was used for removing the bubbles)
11. Run the gel with running buffer at 150 V for 1.5 hours (Amp. 2 is okay)
12. PVDF membrane was cut according to the size of the gel and kept in Methanol (100%) for 20 seconds and then keep in transfer buffer and shake for 20 mins

SDS-PAGE:

13. After running the gel keep SDS-PAGE gel into transfer buffer and shake for 30 mins

Blotting process:

14. Soak the 6 filter papers in transfer buffer

    3 filter paper

    Gel

    Membrane PVDF

    Filter paper

15. 15 V for 2 hours (Keep like that for the whole night-**Overnight**)
16. Next day, take the membrane out after the transfer of the bands from the gel to the membrane
17. Block the membrane with the Blocking one and shake for 1 hour at RT, or 4C for overnight
18. Then use the MYC tag 1^{st} antibody (Monoclonal mouse antibody), 1: 5000 dilution using Solution 1 (0.2 ul of 1^{st} antibody with 1000 ul of Solution 1), 1: 1000 dilution

of Beta actin antibody with Solution 1, Apply the mixture of antibody and solution to the membrane and keep it in the box for 3 hours having some wet tissue on the wither side of the membrane

19. After 3 hours wash the membrane with TBST for 4 time
    i. 5 mins
    ii. 5 mins
    iii. 10 mins
    iv. 10 mins
20. Apply the exact 2nd antibody(Anti-mouse antibody for MYC tag and Protein G HRP for Beta actin), 1:5000 dilution with the Solution 2
21. After application of the 2nd antibody, keep into the box for 1 hour and place the soaked tissue on the either side of the membrane
22. After the end of the time wash the membrane again with TBST for 4 times
    i. 5 mins
    ii. 5 mins
    iii. 10 mins
    iv. 10 mins
23. ECL solution with 1 mL of Solution 1and 1 mL of Solution 2 mixing
24. Wrap the membrane with the film wrap
25. Observe in LAS 3000 (Chemoluminescence, 2 mins manual)

4.2.6 Transfection into BFP stable transformed HEK 293 cells:

BFP stable HEK293 cells (having 50-70% confluency) were transfected with plasmid vectors using Lipofectamin 2000 (Invitrogen), according to the manufacturer's instructions, with 3×10^5 cultured cells per well of a 12-well plate.

About 800 ng of APOBEC 1 deaminase and 700 ng of guide RNA was transfected in each well having total amount of 1 mL optimem (Gibco) along with 4 ul of Lipofectamin 2000 (Invitrogen) per well and kept in to the incubator at 37^0C for 6 hours. After 6 hours, Optimem was replaced by D-MEM (Gibco) to optimize cell growth, and the cells were incubated at 37^0C for 48 hours prior to observation.

4.2.7 Observation of the cells for fluorescence by confocal microscope:

Cells were observed by Laser scanning confocal microscope FV 1000 D (Olympus, Shinjuku-ku, Tokyo, Japan) under the optimized conditions during observation. For very clear image with increased effective resolution and adjusted magnification which helps to observe the exact location of the fluorescence within the cells particularly in a single cell.

4.2.8 RNA extraction and cDNA synthesis from the transfected cells:

Cells of whole dish was harvested after observing under microscope, the total RNA was extracted using TRIzol$^{(R)}$ reagent (Invitrogen) method. Followed by cDNA was synthesized from extracted RNA using a SuperScript™ III First-Strand Synthesis System for RT-PCR (Invitrogen).

4.2.9 Confirmation of the restoration results by PCR-RFLP:

For further confirmation of the restoration, amplified (using Gotaq polymerase, Promega, Fitchburg, WI, USA and PCR machine Gene-Amp PCR system 9700, Applied Biosystems, USA) PCR products were digested with restriction enzyme that differentiated between the edited and unedited DNA sequences. The PCR bands of 324 bp were cut out and purified using the QIAGEN Gel extraction kit (Hilden, Germany). cDNA was quantified using an ND-1000 spectrophotometer. Equal amounts of cDNA

were incubated at 37^0C for 2 hours with BtgI (BioLab) restriction enzyme, which digested the PCR product of 324 bp into two fragments of 201 bp and 123 bp. The BFP was cut (only BFP or APOBEC1 or guide RNA) by the enzyme and the restored GFP remained at the 324 bp position. The remaining portion of the DNA proves the restoration of the C to U (BFP to GFP).

Editing rate by the band intensity: $\frac{\text{Uncut T at 324 bp}}{\text{Uncut T at 324 bp+Cut C at 201 bp + Cut C at 123 bp}}$ X 100 %

4.2.10 Sanger's sequencing:

DNA Sequencing of PCR-amplified products was performed on an applied Biosystems 3130xl Genetic analyzer (Applied Biosystems). PCR amplified fragment was run through 1% Agarose gel and the bands were cut out, frozen and DNA was purified by using QIAGEN Gel extraction kit (Hilden, Germany), the concentration was checked by Nano drop. With the purified DNA, sequencing was done by using the Big Dye Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Technologies) following the manufacturer guide line using the forward and reverse primer for GFP. The raw data of the sequencing were analyzed with the seq scanner 2 software (Applied Biosystems). When the edited and unedited products were presented together, at the target site a dual peak C (unedited) and T (edited) was observed. The rate of editing efficiency was calculated by measuring the area and peak height by ImageJ software (JAVA).

Editing efficiency (Sense) = $\frac{\text{Area of T}}{\text{Area of T + Area of C}}$ X 100%

4.3 RESULTS

I have tried to compare the results with the pol II CMV promoter and pol III U6 promoter. I tried to investigate does the pol III promoter actually make any observable changes in the restoration efficiency in case of the C-to-U editing.

In my previous study I have applied the pol II CMV promoter both in case of the preparation of the deaminase domain construct of APOBEC 1 and also the guideRNA preparation. Here in this study I have tried to prepare the guideRNA under the control of the pol III U6 promoter, and also a single construct having both the pol III and pol II promoters together.

I know that pol III U6 is the best among the other promoters for preparing the short hairpin or sgRNA, as a result I have changed the promoter of the guideRNA to validate the efficacy and compare between the pol II CMV and pol III U6, but I kept the promoter for the deaminase construct same as pol II CMV promoter.

I have observed the results by several experimental and observational methods such as fluorescence observation via the LSM confocal microscope, PCR-RFLP of the extracted cDNA from the positive or restored cells, and Sanger's sequencing confirmation of the restoration of the genetic code.

4.3.1 Western Blot analysis:

From the western blot analysis I have tried to check the binding between the MS2-HB and the APOBEC 1 deaminase. From the protein expression we could see that at 78 kDa level there was a band which proves that the bonding affinity between the MS2-HB an dAPOBEC1 is quite strong. For considering the positive control β-actin was taken and negative control the cell lysate without transfecting the cells with the MS2-HB-APOBEC1 deaminase construct. In case of the Beta actin the band was

observed at 42 kDa and for only cell lysate there was no band observed. But Cellular protein was there which was observed when the beta actin was tested for the protein existence. But when the specific Myc Tag antibody was used it did not show any band as the specific protein was not there, as APOBEC 1 construct was not transfected in this case (Figure 33).

For expressing the MS2-HB-APOBEC 1 protein in to the cells Myc-Tag antibody was used as the construct was prepared in the PCS2-MT vector plasmid.

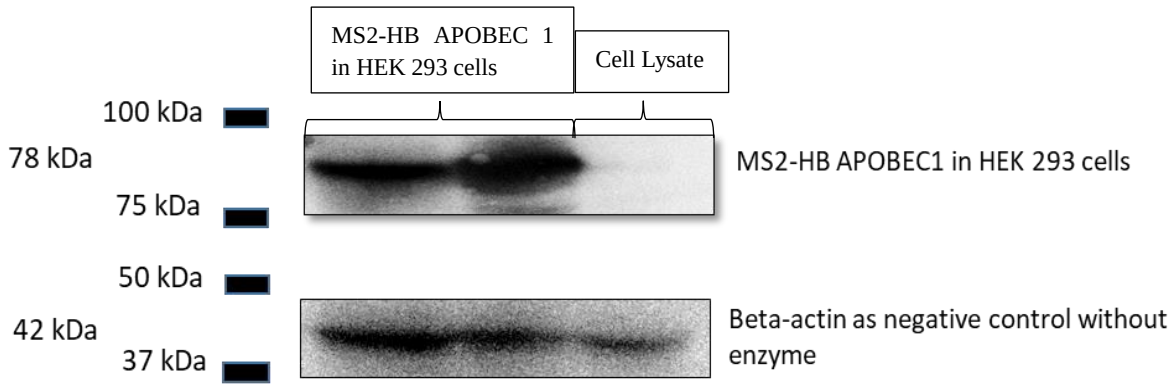


Fig 33. Western blot analysis for checking the protein expression of APOBEC 1 in the HEK 293 cell lines. Beta actin was considered as the negative control without any enzyme which was observed at 42 kDa. The APOBEC1 expression was observed at 78 kDa, Lane 1, 2: MS2-HB-APOBEC1 in HEK 293 cells; Lane 3: Cell Lysate.

4.3.2 Laser Confocal Microscopy for the restored green fluorescence observation:

Unlike the previous study done by our group, in this case I have also observed the restored green fluorescence along with the positive control (wild type GFP) and negative controls (either only the deaminase or the guideRNA) by using LSM confocal microscope for more clearer images.

I did the experiment with the CMV controlled guideRNA, U6 controlled guideRNA and also the single construct, having the deaminase and the guideRNA in the same construct

under the control of the two different promoters; for APOBEC 1 deaminase pol II CMV promoter and for guideRNA pol III U6 promoter (Figure 34).

I observed that when the CMV controlled APOBEC 1 and guideRNA was transfected into the BFP stably transformed cells, the intensity of the restored GFP was a little lesser than that of the CMV controlled APOBEC 1 and U6 controlled guideRNA combination or even transfected with single construct. I calculated the light intensity in all the cases and I have found that in case of the guide which was prepared under the control of the CMV promoter, 21% was the restored percentage among all the cells, whereas in case of the guideRNA prepared under the control of the U6 promoter, it was 44% and for single construct (APOBEC 1 and guideRNA in the same construct) the restoration percentage was 51% (Figure 35).

Before the application of the single construct for the purpose of editing and also a set of comparable level of different guides, I standardized the molecular weight between all the constructs and then did the transfection because the molecular weight of the single construct is not equivalent to deaminase and guideRNA irrespective of promoter. Due to combination of the APOBEC 1 and guideRNA at a single construct the molecular weight is much higher than the individual constructs of individual APOBEC 1 deaminase and guideRNA.

I also kept the wild type GFP as a positive control whereas the transfection with the only one construct either deaminase or guideRNA (in case of U6 promoter as for CMV it was tested in the previous study as well) was done to be considered as a negative control, where I found that one factor either deaminase or guideRNA alone has no capacity to do the restoration/editing of the genetic code. Only when both the factors are transfected together or remained in a single construct only then the editing of the C to U

for the restoration of the genetic code could be possible.

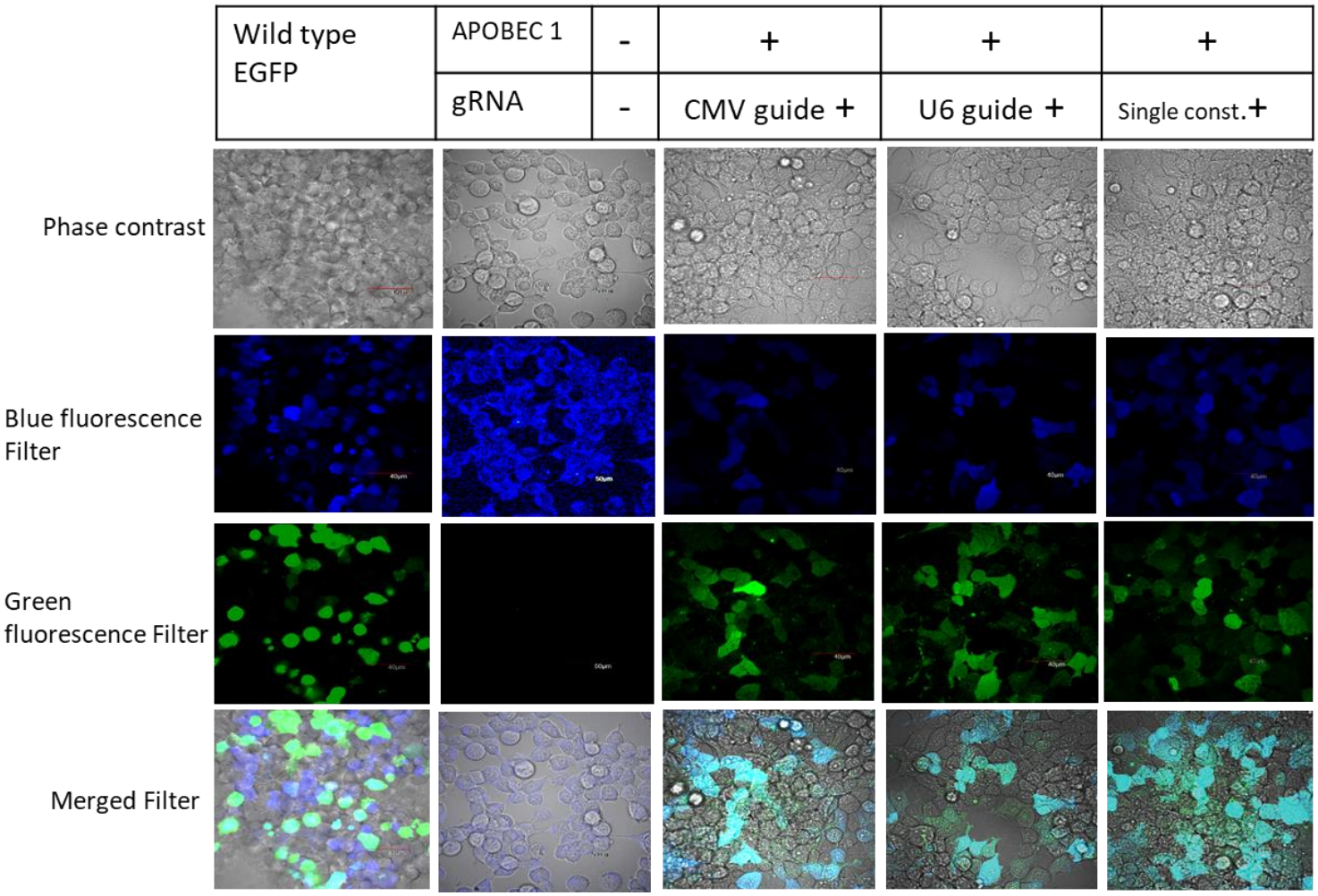


Fig 34. Confocal microscopic images for observation of the fluorescence after restoration. There are five panels where wild type EGFP is the positive control where there was green fluorescence and only BFP cells without enzyme or guideRNA is the negative control where there was only blue fluorescence but no green fluorescence. In case of the CMV containing guideRNA and APOBEC 1 there was appearance of the green fluorescence due to the restoration of the genetic code from C to U by editing and same happened in case of the U6 promoter and single construct. But the green fluorescence intensity differs from each other. Although BFP was also observed in case of the restored samples as well because its not possible to restore 100% of target mRNA. But if from the negative control the BFP is compared with the BFP in restored samples the intensity was also reduced in case of the restored samples.

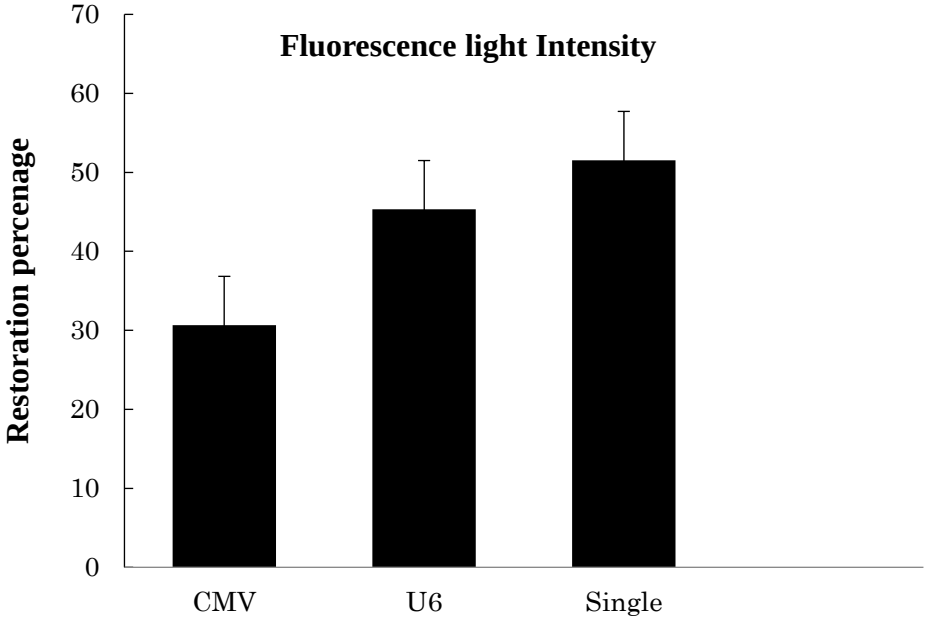


Fig 35. Restoration percentage according to Light intensity of BFP and GFP. In case of the CMV the restoration percentage is 30.65% whereas in case of U6 and Single construct the percentage is 45.31% and 51.53%, respectively. U6 gives better restoration percentage than CMV but Single construct prepared with the combination of the CMV promoter containing APOBEC1 and U6 promoter containing guideRNA was the best among the three. For all data the statistical analysis (mean±SEM) was done, where n=3.

4.3.3 PCR-RFLP confirmation for the restoration of the genetic code:

For the PCR-RFLP confirmation of the restoration, the PCR amplified cDNA synthesized from the positive cells were digested with the BtgI restriction enzyme. According to the sequence preference the restriction enzyme could cut only the BFP (CCC) but neither the wild type GFP (TCC) nor the restored GFP.

As a result the BFP was completely cut at 201 and 123 bp by the restriction enzyme, the total length of the amplified fragment was 324 bp. But in case of the wild type of GFP and restored GFP there was a remained band at 324 bp. But I could also see the small bands at 201 and 123 bp because the transfection was done in BFP stably transformed HEK 293 cells. Though I could see from the confocal microscopy that in the dishes, where the transfection was done with the two factors the intensity of the blue

fluorescence light reduced among those which were still not edited because the 100% editing is not possible.

I tried to calculate the restoration percentage from the band intensity of the RFLP bands. From the calculation I have found, different rate of restoration from different concentrations of the deaminase and the guideRNA combination such as APOBEC1+guideRNA: 400+200 ng/ well, 400+300 ng/ well, 400+400 ng/ well and 600+500 ng/ well respectively. In all the cases we have found that the restoration rate has been increased according to the increase of either the guideRNA or the deaminase. Similarly for the U6 promoter, APOBEC1+guideRNA: 400+200 ng/ well, 400+300 ng/ well, 400+400 ng/ well and 600+500 ng/ well, concentrations were used for the transfection, the speciality of U6 is that in this case APOBEC1 is under CMV promoter but guideRNA is under U6 promoter. Similar results were observed even in case of the U6 promoter, that is with the increase of the concentration of either APOBEC1 deaminase or guideRNA the restoration rate was increased. The restoration percentages were as follows:

Table 6. Different restoration percentage (average) in case of CMV and U6 promoter in relation to the increase of the concentration of the deaminase APOBEC1 or guideRNA

Different concentration	**Editing rate (hCMV-4 samples and U6-5 samples)**
hCMV-400+200	0.20
hCMV-400+300	0.21
hCMV-400+400	0.23
hCMV-600+500	0.27
U6-400+200	0.21
U6-400+300	0.27
U6-400+400	0.24
U6-500+600	0.26
U6-600+500	0.34

If I compare between the CMV and U6 promoters, from the calculation of the restoration percentage I could understand that with the same concentration, the restoration percentage of CMV and U6 are different. Such as when, APOBEC1 concentration 400 ng/ well and guideRNA 200 ng/ul under the control of the CMV pol II promoter the restoration percentage was 20% but with the same concentration under the U6 pol III promoter the percentage was 21%. Though it was not so big difference but with the change of the concentration the differences also increased, such as when 600ng/ well+500ng/ well concentrations were used the restoration percentage for the CMV was 27% whereas for U6 34% which showed a significant difference (Figure 36). All though there is a concern that whether with the increased concentration of the guideRNA and APOBEC 1 deaminase, the off target effect may also be affected or not.

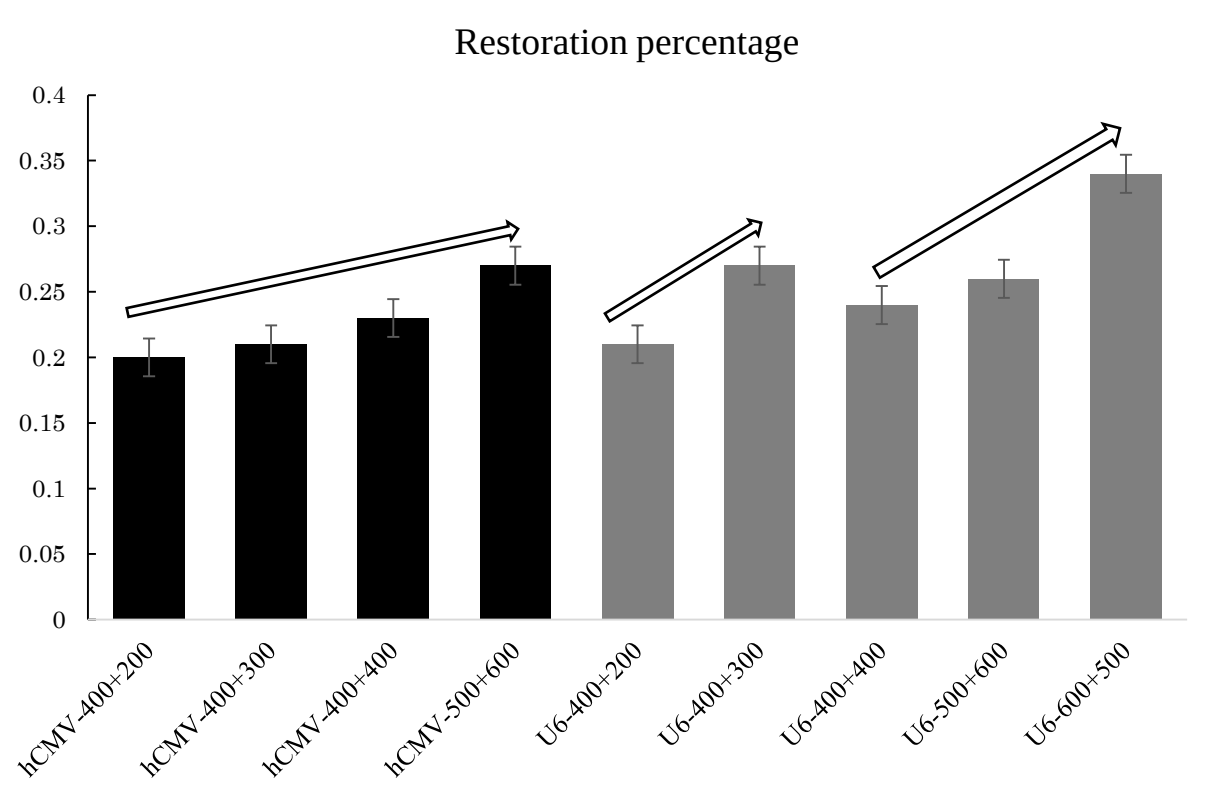


Fig 36. Comparison between CMV and U6 promoter efficiency in respect to restoration percentage at different concentrations during the time of transfection. For all data the statistical analysis (mean±SEM) was done, where n=3.

4.3.4 Sanger's sequencing confirmation of the genetic code restoration:

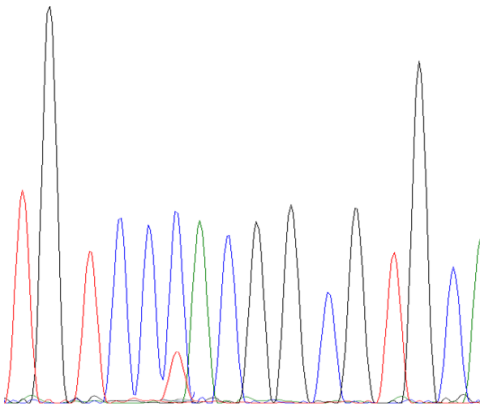
For the further confirmation of the restoration of genetic code by RNA editing from C-to-U and also to check the off target effects, as a vital evidence Sanger's sequencing was done with the forward or sense primer. When the genetic code was restored there could be seen a dual peak at the target nucleotide, one of those was for the mutated and another one was for the restored. In case of the restored GFP the editing happened from CCC to CCT, as a result I could see the dual peaks of C and T at the position of target nucleotide of C (In case where both the APOBEC 1 and guide RNA was transfected)

(Figure 37).

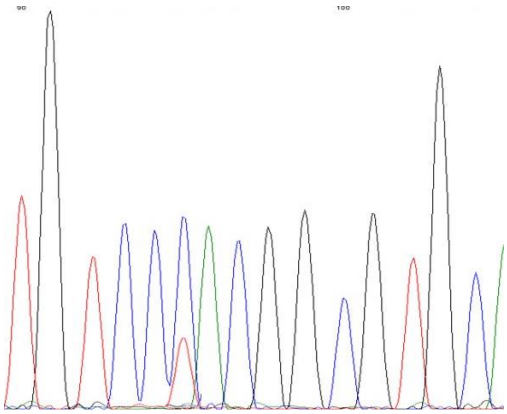
If I compare between the three different types of approaches along with guideRNAs with different promoters and the single construct based on the editing efficiency calculated from the peak height and peak area of the Sanger's sequencing, I could state that in case of the CMV promoter controlled deaminase and guideRNA the editing efficiency was 21% whereas in case of the CMV controlled deaminase and U6 controlled guideRNA (as the individual plasmid construct) the editing efficiency was 39.4%. I transfected the 600 ng/well of deaminase and 500 ng/ well of guideRNA in to the BFP stably transformed HEK 293 cells (Figure 38).

Then if I compare the editing efficiency with the Single construct, I obtained that the editing efficiency was 41.65% which was higher than that of the combination of CMV controlled deaminase and U6 controlled guideRNA in two different plasmids, where the two factors were co-transfected. It could be as the Single construct was also prepared with the combination of the two different promoters CMV for deaminase and U6 for guideRNA, but in a single plasmid. So from the pol II to pol III promoter change made a vital increase in the editing efficiency, whereas the single construct comparing to individual had higher editing efficiency.

CMV promoter containing gRNA



U6 promoter containing gRNA



Single construct containing construct

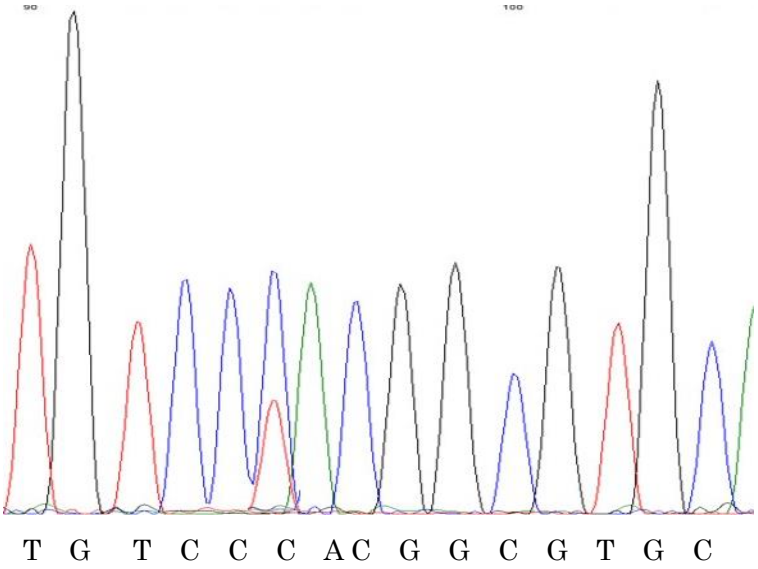
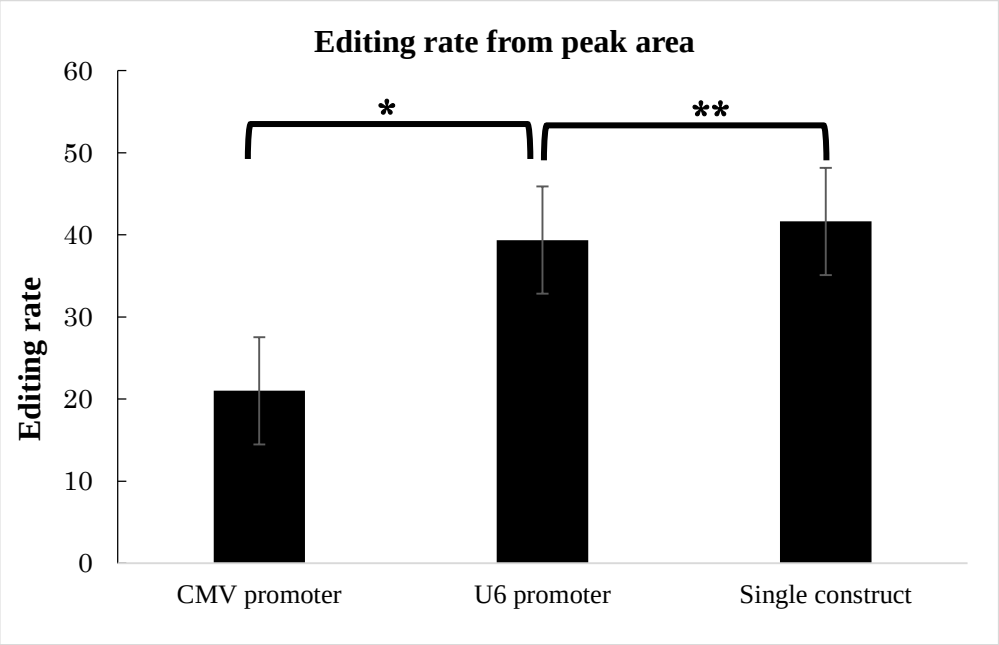


Fig 37. Comparative sequence result of the different promoters (U6 and CMV) and Single construct (prepared from the U6 guide RNA and CMV promoter containing APOBEC 1)



*=the p value is significant at 95% confidence interval, as p< 0.01

**= the p value is significant at 95% confidence interval as p<0.05

Fig 38. Editing rate from the peak area of the different promoters (U6 and CMV) and Single construct (prepared from the U6 guide RNA and CMV promoter containing APOBEC1) of the sequence result. Where, CMV promoter shows 21.02% and in case of U6 and single construct 39.37% & 41.65% respectively, student t-test was done, where n=3.

``

4.4 DISCUSSION

In this study I have tried to make a difference between the pol II and pol III promoters and also the single construct in respect to editing efficiency from C-to-U restoration. From the experimental data I have found that in case of the CMV promoter controlled process of RNA editing where both the deaminase and guideRNA constructs were prepared under the control of the pol II CMV promoter, the editing efficiency was lesser comparing to the U6 promoter containing guideRNA or in single construct having combined approach of CMV in deaminase and U6 promoter in guideRNA construct. Unlike the CMV controlled editing approach, U6 promoter containing guide RNA leading editing approach also showed that, the one factor either deaminase or guideRNA could not restore the genetic code by itself. Rather when both the factors were transfected together only then the fluorescence was observed as a result the restoration of the genetic code had happened. In case of the single construct approach the transfection factor became one to be applied in to the BFP stably transformed HEK 293 cell lines, which reduces the cell stress due to the less amount of factors to be transfected.

This is the first time where the U6 promoter has been used along with the MS2 system for the sake of RNA editing and I have successfully achieved that, even first time as a single construct as well. Previously the combination of U6 and CMV promoters have been used where the CMV was placed upstream of the U6, following those previous studies I have also used the combination of the CMV and U6 promoters where CMV has been placed upstream of the U6, but U6 promoter was being attached with the guideRNA and the MS2-stem loop which is unique in this current editing approach. The combination of the CMV promoter and U6 promoter along with the guideRNA and

MS2 system has been used for the first time for restoration of C to U genetic code *in vivo*.

Montiel–Gonzalez et al., (2016) **(**20**)** used the Lambda N system for the site directed RNA editing along with the ADAR deaminase, where they prepared the guideRNA under the control of pol III U6 promoter. But in this study I have used the MS2 system, which is more efficient and has been used more commonly in RNA studies than the Lambda N system; in respect of the editing efficiency MS2 is stronger than Lambda N system and efficient. The bonding affinity towards the hairpin is 10^{-8} M for the Lambda N system, whereas the affinity towards the AUUA stem loop sequence is 10^{-9} M for MS2, which improves to 10^{-10} when the AUCA stem loop sequence is present. This explains why MS2 is stronger than the Lambda N system (20). And, MS2 along with the U6 promoter will definitely increase the efficiency of the enzymatic system for the genetic code restoration.

Concerning the restoration of the genetic code with the developed artificial enzyme system, in my previous study only considering the CMV promoter controlled approach I found that when both the factors were transfected together for the purpose of restoration only then I could observe the green fluorescence but only one factor either deaminase or guideRNA, irrespective of promoter could not do the editing. The same result was also found in case of the U6 promoter controlled guideRNA and CMV controlled deaminase approach as a separate plasmid construct, which has indicated that the other factors relevant to cells are not responsible for the editing. Further optimization of the system, including the guideRNA sequence, the usability of the MS2 moiety with the enzyme, and the relative proportions of the two factors, has to be performed in future work. Appropriate concentration of the deaminase and the reporter substrate can eliminate the

minor auto editing which may cause off target effects.

From the LSM confocal microscopy result I could observe that the intensity of the restored green fluorescence was different in case of different approaches. In case of the single construct it was the highest in comparison with the U6 controlled guideRNA or CMV controlled guideRNA approach along with the APOBEC1 deaminase which is controlled by the CMV promoter. But if I compare between the U6 controlled guideRNA and CMV controlled guideRNA approach along with the CMV controlled deaminase and then I could find that in case of U6 controlled guideRNA, the intensity was around 45.31% whereas in case of CMV it was 30.65% and in case of the Single construct (with the U6 and CMV combined promoter) it was 51.53%.

So the change of the promoter increased the efficiency of the editing of the developed artificial deaminase system which supported the findings of the Su Mi kim et al., (2010), who compared the effectiveness of pol II and pol III promoters and found that pol III had better efficacy than pol II but their experiment was gene knockdown oriented (21). Regarding the effectiveness of the single construct with a combination of the CMV and U6 promoters, the experimental data supported the findings of Jianguo Su et al., (2008) (22), who found that from the hybrid construct produced when the CMV promoter or enhancer is placed to the immediate upstream of the U6 promoter, the silencing efficiency by the shRNA increased. The hybrid construct of CMV deaminase and U6 guideRNA where the CMV was placed immediately to the upstream of the U6 promoter. I also observed from the experimental data, both in case of the fluorescence intensity and also from the PCR-RFLP band intensity the restoration percentage in case of the single construct was higher than any other approach.

I have also observed from the PCR-RFLP (band intensity) data that with the increase of

the concentration of the deaminase or the guideRNA, the restoration percentage had also increased. But the main concern in this regard was the off target effect. Because from one of the previous experiments of our group I found that if I increase the concentration of deaminase to be transfected per well then there is a possibility that the editing rate may increase but at the same time it may increase the off target effects as well. So in the future work I will consider this particular issue of off target effect which I haven't done yet. From the sequencing data of the Sanger's sequencing, I could see the dual peak appeared only when the two factors were transfected together or after the transfection of the single construct.

I also calculated the editing efficiency from the peak height of the Sanger's sequencing data. After the calculation of the efficiency I found that in case of the CMV controlled approach the rate was 21.02% whereas in case of the U6 controlled and single construct the restoration rate was 39.37% and 41.65%, respectively, which also supported the data of Su mi Kim et al., and Jianguo et al., (21, 22).

From the respective experimental data I have surely observed that pol III worked better than the pol II promoter for the purpose of RNA editing whereas the single construct (a combination of the CMV and U6 promoter) worked even better than the CMV and U6 combination but in two different plasmid vectors. If I compare the editing efficiency rates achieved between the CMV controlled guideRNA approach and U6 controlled guideRNA approach along with the CMV controlled APOBEC 1 deaminase the p Value is 0.0044, in case of U6 controlled guide approach the editing efficiency has significantly increased than that of CMV approach. Again regarding the comparison between Single construct and the U6 promoter p value is 0.0007 which is highly significant, so in case of the single construct the editing efficiency has significantly.

4.5 CONCLUSION

This study demonstrated that the comparative editing efficiency between the CMV promoter and U6 promoter, the U6 promoter which is polymerase III has given a better editing efficiency than that of the polymerase II CMV promoter. When the comparison between the U6 promoter and the Single construct prepared by a combination of polymerase II CMV promoter with the APOBEC 1 deaminase & polymerase III U6 promoter with the guideRNA, in case of the single construct the editing efficiency was higher than that of the U6 promoter. So among the three types of promoter constructs (only CMV, only U6 and combination of CMV and U6) the single construct gave the best editing rate. This single construct approach might be very helpful during the animal model transfection via the viral vector. The success of this enzymatic approach of RNA editing will open a new type of therapeutic approach for the restoration of the genetic code in case of the diseases caused by T-to-C point mutation.

Moreover, the approach that uses multiple shRNAs driven by the combination of Pol II and Pol III promoters, is probably a more desirable strategy for antiviral efficiency without cell toxicity. However, if I use the approach of the combined pol II CMV and pol III U6 promoters in natural host models, the toxicity might be negligible.

4.6 REFERENCES

1. Heidenreich M., Zhang F. Applications of CRISPR-Cas systems in neuroscience. Nat Rev Neurosci., 17(1): 36–44, (2016).
2. Kim Y.G., Cha J., Chandrasegaran S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proceedings of the National Academy of Sciences. 93(3): 1156–1160, (1996).
3. Mali P., Yang L., Esvelt K.M., Aach J, Guell M., DiCarlo J.E., Norville J.E., and Church G.M. RNA-guided human genome engineering via Cas9. Science, 339(6121): 823–826, (2013).
4. Porteus M.H., and Carroll D. Gene targeting using zinc finger nucleases. Nat Biotechnol, 23(8): 967–973, (2005).
5. Zhang F., Le C., Simona L., Sriram K., George M.C., Paol A. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. Nat Biotechnol, 29(2): 149–153, (2011).
6. Thomas G., Shannon J.S., Sai-lan S., and Jia L. Genome editing technologies: Principles and applications. Cold spring Harbor Perspectives in Biology, 8:a023754, (2016).
7. Aisling M.F., and Andrew R.G. Adenosine deaminase deficiency: A review. Orphanet: Journal of Rare diseases, 13: 65, (2018).
8. Lane D.A., Kunz G., Olds R.J., and Thein S.L. Molecular genetics of antithrombin deficiency". Blood Rev., 10 (2): 59–74, (1996).
9. Johansson H.E., Liljas L., and Uhlenbeck O.C. RNA recognition by the MS2 phage coat protein. Sem Virol, 8(3): 176–185, (1997).
10. Liqun L., Chen H., Xue W., Yang B., Hu B., Wei J., Wang L., Cui Y., Li W., Wang J.,

Yan L., Shang W., Gao J., Sha J., Zhuang M., Huang X., Shen B., Yang L., and Chen J. APOBEC3 induces mutations during repair of CRISPR–Cas9-generated DNA breaks. Nature Structural & Molecular Biology, 25: 45–52, (2018).

11. Hsu P.D., Lander E.S., and Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. Cell, 157(6): 1262–78, (2014).
12. 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).
13. Bhakta S., Azad M.T.A., and Tsukahara T. Genetic code restoration by artificial RNA editing of Ochre stop codon with ADAR1deaminase. Protein Engineering, Design & Selection, 31(12):1-8, (2018).
14. Brenz V.M.S., Weber P., Mayer C., Graf C., Refojo D., Kuhn R., Grummt I., and Lutz B. Development of a species specific RNA polymerase I-based shRNA expression vector. Nucleic Acids Res., 35(2):e10, (2007).
15. Li X., Jiang D.H., Yong K., and Zhang D.B. Varied transcriptional efficiencies of multiple Arabidopsis U6 small nuclear RNA genes, J. Integr. Plant Biol., 49(2):222-229, (2007).
16. Belhaj K., Chaparro-Garcia A., Kamoun S., and Nekrasov V. Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas system. Plant Methods., 9: 39, (2013).
17. Cong L., Ran F.A., Cox D., et al.. Multiplex genome engineering using CRISPR/Cas systems. J. Science., 339(6121): 819-823, (2013).
18. Su J., Zhu Z., Xiong F., and Wang Y. Hybrid Cytomegalovirus-U6 Promoter-based Plasmid Vectors Improve Efficiency of RNA Interference in Zebrafish. Marine

Biotechnol., 10(5):511–517, (2008).

19. Satheesh V., Zhang H., Wang X., and Lei M. Precise Editing of Plant Genomese Prospects and Challenges. Seminars in Cell & Developmental Biology. **In press** (2019).
20. Long L., Guo D.D., Gao W., Yang W.W., Hou L.P., Ma X.N., Miao Y.C., Botella J.R., and Song C.P. Optimization of CRISPR/Cas9 genome editing in cotton by improved sgRNA expression. Plant Methods, 14 (85): 1-9, (2018).
21. Qi W., Zhu T., Tian Z., Li C., Zhang W., and Song R. High-efficiency CRISPR/Cas9 multiplex gene editing using the glycine tRNA-processing system-based strategy in maize. BMC Biotechnol., 16 (58): 1-8, (2016).
22. Zhang Z., Hua L., Gupta A., Tricoli D., Edwards K.J., Yang B., and Li W. Development of an Agrobacterium-delivered CRISPR/Cas9 system for wheat genome editing. Plant Biotechnology Journal., 17: 1623-1635, (2019).
23. Sun X., Hu Z., Chen R., Jiang Q., Song G., Zhang H., and Xi Y. Targeted mutagenesis in soybean using the CRISPR-Cas9 system. Sci. Rep., 5:10342, (2015).
24. Song J., Pang S., Lua Y., Chiu R. Poly (U) and polyadenylation termination signals are interchangeable for terminating the expression of shRNA from a pol II promoter. Biochem Biophys Res Commun., 323:573–578, (2004).
25. Su J., Zhu Z., Wang Y., Xiong F., and Zou J. The cytomegalovirus promoter-driven short hairpin RNA constructs mediate effective RNA interference in zebrafish in vivo. Mar Biotechnol, 10(3): 262–269, (2007).
26. Xia H., Mao Q., Paulson H.L., and Davidson B.L. siRNA-mediated gene silencing in vitro and in vivo. Nat Biotechnol, 20:1006–1010, (2002).
27. Xia X.G., Zhou H., Ding H., Affar E.B., Shi Y., and Xu Z. An enhanced U6

promoter for synthesis of short hairpin RNA. Nucleic Acids Res 31:e100, (2003).

28. Su J., Zhu Z., Xiong F., and Wang Y. Hybrid Cytomegalovirus-U6 Promoter-based Plasmid Vectors Improve Efficiency of RNA Interference in Zebrafish. Mar Biotechnol, 10:511–517, (2008).
29. Brummelkamp T.R., Bernards R., and Agami R.A. system for stable expression of short interfering RNAs in mammalian cells. Science, 296:550–553, (2002).
30. Jacque J.M., Triques K., and Stevenson M. Modulation of HIV-1 replication by RNA interference. Nature, 418:435–438, (2002).
31. Montiel-Gonz´alez M.F., Vallecillo-Viejo I.C., and Rosenthal J.J.C. An efficient system for selectively altering genetic information within mRNAs. Nucleic Acids Research, 44 (21): e157, (2016). doi: 10.1093/nar/gkw738
32. Kim S.M., Lee K.N., Lee S.J., Ko Y.J., Lee H.S., Kweon C.H., Kim H.S., and Park J.H. Multiple shRNAs driven by U6 and CMV promoter enhances efficiency of antiviral effects against foot-and-mouth disease virus. Antiviral Research, 87 (3): 307-317, (2010).
33. Hassani Z., François J.C., Alfama G., Dubois G.M., Paris M., Giovannangeli C., and Demeneix B.A. A hybrid CMV-H1 construct improves efficiency of PEI-delivered shRNA in the mouse brain. Nucleic Acids Research, 35(9): e65, (2007).
34. Paul C.P. Effective expression of small interfering RNA in human cells. Nat. Biotechnol., 20:505–508, (2002).
35. Xia X.G. An enhanced U6 promoter for synthesis of short hairpin RNA. Nucleic Acids Res.; 31:e100, (2003).
36. Ong S.T., Li F., Du J., Tan Y.W., and Wang S. Hybrid cytomegalovirus enhancer-h1 promoter-based plasmid and baculovirus vectors mediate effective RNA interference.

Hum. Gene. Ther. 16:1404-1412, (2006).

37. Matthess Y. Conditional inhibition of cancer cell proliferation by tetracycline-responsive, H1 promoter-driven silencing of PLK1. Oncogene, 24:2973-2980, (2005).
38. Thompson J.D. Improved accumulation and activity of ribozymes expressed from a tRNA-based RNA polymerase III promoter. Nucleic Acids Res., 23:2259–2268, (1995).
39. Ilves H. Retroviral vectors designed for targeted expression of RNA polymerase III-driven transcripts: a comparative study. Gene, 171:203–208, (1996).
40. Boden D. Promoter choice affects the potency of HIV-1 specific RNA interference. Nucleic Acids Res., 31:5033–5038, (2003).
41. Hassani Z. A hybrid CMV-H1 construct improves efficiency of PEI-delivered shRNA in the mouse brain. Nucleic Acids Res., 35:e65, (2007).
42. Luyen T.V., Thanh T.K.N., Shafiul A., Takashi S., Kenzo F., Hitoshi S., and Toshifumi T. Changing Blue Fluorescent Protein to Green Fluorescent Protein Using Chemical RNA Editing as a Novel Strategy in Genetic Restoration. Chem Biol Drug Des, 86: 1242–1252, (2015).
43. Luyen T.V., and Toshifumi T. C-to-U editing and site-directed RNA editing for the correction of genetic mutations. BioScience Trends Advance Publication, 11(3): 243-253, (2017).

CHAPTER V

Restoration of the genetic code from C-to-U in real model mouse (Macular Mouse)

5.1 INTRODUCTION

For the application of the developed artificial RNA editing system the macular mouse having Menkes kinky disease was chosen. Menkes kinky hair is a recessive hereditary disease in humans, linked in the X-chromosome and caused by defective copper metabolism (1, 2), particularly mutations in genes coding for the copper-transport protein. Development impediment, pericular hair, and neurological symptoms attributable to cerebral and cerebellar degeneration are the characteristics of the sickness (5). Additional signs and indications include feeble muscle tone (hypotonia), drooping facial features, seizures, formative deferral and psychological disability. Kids born with Menkes syndrome typically start to develop symptoms during infancy and usually do not live past age three. Early treatment with copper may improve the prognosis in some affected individuals. In uncommon cases, symptoms start to show later in adolescence (22).

The gene of Menkes disease was isolated by a positional cloning method (6, 7, 8). The protein encoded by the gene is copper-transporting P-type ATPase (ATP7A). The ATP7A gene provides directions for making a protein that is significant for controlling copper levels in the body. Copper is fundamental component for many cellular functions, but it is toxic when present in excessive quantities. Mutations in the ATP7A gene result in poor dispersion of copper ion to the cells of the body (4). Copper aggregates in certain tissues, such as the small intestine and kidneys, while the brain and other tissues have surprisingly low levels of copper. The decreased supply of copper can decrease the activity of various copper-containing enzymes that are essential for the structure and function of bone, skin, hair, blood vessels, and the nervous system. The signs and symptoms of Menkes syndrome and occipital horn disorders are caused by the

diminished action of these copper-containing enzymes (11).

The mottled series of mutations on the mouse X-Chromosome (Chr) bring about faulty copper metabolism, and mice with these mutations are considered to be models of human Menkes syndrome (Levinson et al. 1994; Mercer et al. 1994). The macular mouse is a mutant strain that emerged in C3Hf stock and has an alternate mutant allele (the old symbol was Mo^{ml}, and the new symbol is ATP7A $Mo^{\sim ml}$) on the mottled locus. The macular mouse is thought to be an ideal model for human Menkes disease.

The macular mouse was relied upon to show differentiation in the ATP7A gene as in the dappled and the blotchy mice. However, Northern and Southern investigations of the ATP7A gene have revealed no gross variations from the normal to the macular mouse (9, 10). An in situ hybridization study revealed that the mRNA for the ATP7A was expressed in macular mouse brain, in the same way as in normal mice (23). In this manner, small nucleotide modification was associated as a reason of the deformities in the macular mouse.

Later on the researchers investigated that the DNA sequence of the ATP7A gene of the macular mouse, where they observed a point mutation that outcomes in non-conservative amino acid substitution, serine to proline, was recognized. The mutation was first observed in the cDNA prepared from the RNA of spleen from the test animals. The mutation was additionally affirmed by investigation of genomic DNA from macular mouse, using the polymorphic BamHI site decimated by the mutation (2). Sanger's sequencing analysis of the synthesized cDNA demonstrated the presence of the T-to-C mutation from the samples of not only spleen but also liver, collected from the 7 days old hemizygous male mouse.

cDNA sequence of the copper-transporting P-type ATPase (ATP7A) gene of the macular

mouse has a point mutation (T-to-C) that results in substitution of proline for serine in a putative eighth transmembrane domain of the ATP7A was identified.

The macular mutant mouse has some phenotypic articulation. The hemizygotes start to show white fur color and curly whiskers around day 3 of postnatal, then seizures and ataxia around day 8, while the typical littermates has not. The hemizygotes also increase weight gradually from birth to day 9, but then showed weight reduction and died around day 15 with extreme emaciation. These clinical features resembled those in Menkes kinky hair disease (13).

There is a physiological feature in case of the Macular mouse. Changes of body weight of 7 to 10 days old mice are observed. No significant difference in body weight could be observed through days 7 to 14 in macular males (Ml/y). Although the body weight of littermates (+/y) at 7 days of age was almost the same as that of Ml/y. the former gained more weight than the latter at 10 days of age, and no significant increase was observed after 10 days . So after 10 days the hemizygous macular male mice stated to lose weight (14, 16).

In this study I have tried to restore the genetic code from C to U in the macular mouse derived fibroblast cells with the guide RNA along the MS2 stem loop under the control of the pol III U6 promoter, and APOBEC 1 deaminase with the MS2-coat protein under the pol II CMV promoter.

5.2 MATERIALS AND METHODS

5.2.1 Rearing of Macular mouse ($Mo^{\sim ml}$):

Macular mice were kindly given by Professor Dr. Hiroko Kodama and Ms. Yoko Hiro, Teikyo University.

The mutant mice have been propagated by mating heterozygous females (Ml/+) with normal males (+/y) and maintained in an animal house with constant temperature (22^0C) and a 12-h light/12-h dark cycle. Tap water and a commercial stock diet (Japan CLEA CE-2) were provided ad libitum. Hemizygous male mice (M1/y) were used as the experimental animals and their normal littermates (+/y) as the controls. The mice were weighed on days 7, 10 and 14 postpartum.

5.2.2 Collection of samples:

The mice were weighed and sacrificed on days 7 postpartum by cervical dislocation. The liver and spleen were removed immediately, weighed and frozen at -80^0C until processed for further experiments.

5.2.3 RNA extraction and followed by cDNA synthesis from the collected samples:

The frozen liver and spleen samples were grinded as much as possible by the gel grinder then TRIzol (Invitrogen) was added to the sample (500 ul / tube), after that Chloroform (100 ul/ tube) was added to the tube. Vortexing was done for 5 mins at full speed. Afterwards centrifugation was done at 12000xg for 15 mins. The supernatant was collected into a new tube, then 500 of isopropanol was added to each tube and a slight vortexing was done. Followed by centrifugation was done at 12000xg for 10 mins. After this step the RNA at the bottom of the tube was seen. Then the supernatant was removed and 500 ul of 70% ethanol was added gently to each tube and final centrifugation was one at 8000Xg for 5 mins. Then the supernatant was removed and the tubes were air

dried. The elution of the RNA was done with 20 ul TE buffer. Concentration of RNA was measured by using Nano Drop. 500 ng of RNA was taken for the purpose of cDNA synthesis. Superscript(R) III (Invitrogen) was for the RT-PCR. The protocol provided by the manufacturer was followed and finally the concentration of the cDNA was measured by using the Nano Drop.

5.2.4 Identification of the Target mutated C in the ATP7A gene:

For the identification of the Atp7a gene from the liver and spleen sample of the Macular mouse model, PCR amplification of the synthesized cDNA was done by using the specific primers. For the amplification of the Atp7a gene the primers those were used are: Fw primer: CTGGATGTTGTGGCAAGTATTGAC; Reve primer: GCTGTTCAGGGAGCGCTTG. The total length of the amplified fragment was 466 bp.

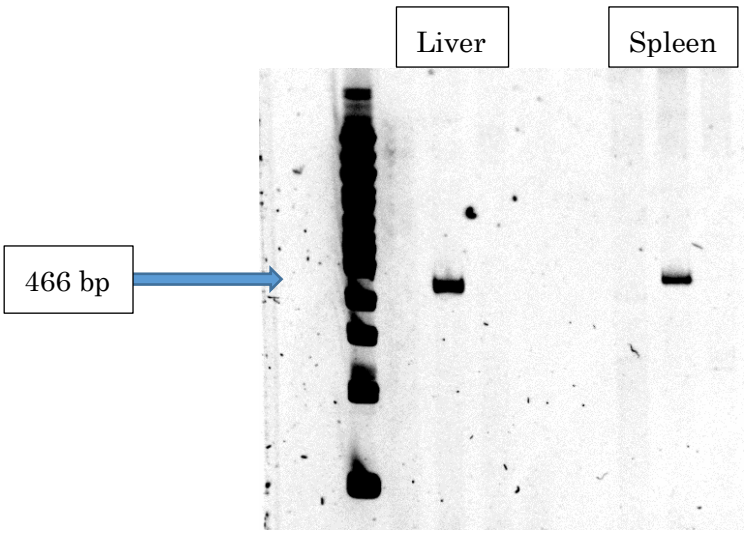


Fig 39. PCR amplification of the ATP7A gene from the cDNA synthesized from the Macular mouse liver and spleen

5.2.5 Sequence confirmation of the mutation (T-to-C) in ATP7A gene:

After the PCR amplification of the fragment of ATP7A gene, the band was cut and DNA

was purified by using the Gel purification kit (QIAGEN gel purification kit, Hilden, Germany). Then the purified sample was sequenced by the Sanger's sequencing method. The sequence data was analyzed by Applied Biosynthesis software and the exact location of the mutation was identified.

Liver:

GCTCTAATTTATAATCTGGTTGGAATTCCCATCGCTGCTGGAGTTTTCCTGCCCATC
GGTTTGGTTTTACAACCCTGGATGGGACCCGCAGCAATGGCCGCTTCATCTGTCTC
TGTAGTCCTTTCTTCCCTTTTCCTCAAGCTTTACAGGAAACCTACGTATGACAATTA
TGAGTTGCATCCCCGGAGCCACACAGGACAGAGGAGTCCTTCAGAAATCAGTGTT
CACGTTGGAATAGATGACACCTCAAGAAATTCTCCAAGGCTGGGTTTGCTGGACCG
GATTGTCAACTATAGCAGAGCCTCCATAAATTCACTGCTGTCTGACAAGCGCTCCC
TGAACAGC

Spleen:

AATAATATATAATCTGGTTGGAATTCCCATCGCTGCTGGAGTTATCCTGCCCATCGG
TTTGGTTTTACAACCCTGGATGGGACCCGCAGCAATGGCCGCTTCATCAGTCTCTG
TAGTCCTTTCTTCCCTTTTCCTCAAGCTTTACAGGAAACCTACGTATGACAATTATG
AGTTGCATCCCCGGAGCCACACAGGACAGAGGAGTCCTTCAGAAATCAGTGTTCA
CGTTGGAATAGATGACACCTCAAGAAATTCTCCAAGGCTGGGTTTGCTGGACCGGA
TTGTCAAC

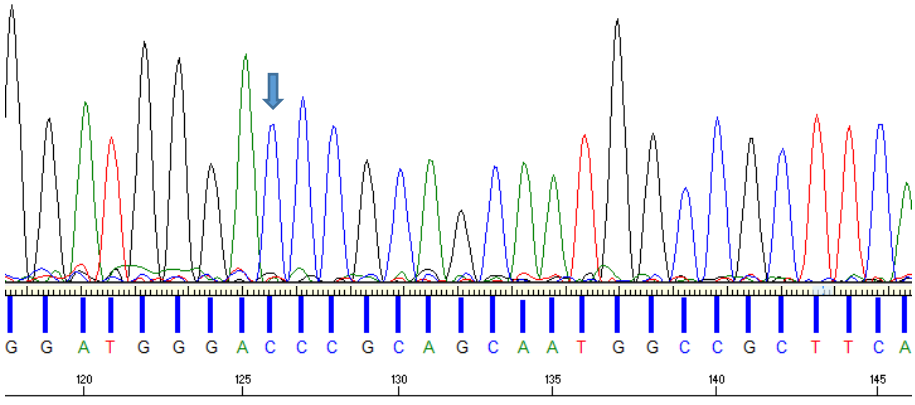


Fig 40. Sequence confirmation of the T to C mutation in ATP7A gene, from the collected Liver and spleen samples.

5.2.6 APOBEC 1 deaminase enzyme preparation:

To enable the enzymes directing to a BFP target codon of interest, the deaminase domain of APOBEC1 was cloned downstream of MS2 in pCS2+MT using the XhoI and XbaI (Takara, Shiga, Japan) restriction enzymes to yield pCS2+MT-MS2HB-APOBEC 1, by PCR amplification from HEK 293 cell line and HeLa cells, the forward and reverse primers consisting of particular restriction sites (XhoI catalytic APOBEC1 Fw: tccactcgagatgccctgggagtttgacgtctt, XbaI catalytic APOBEC1 Rv1: acggtctagattaagggtgccgactcagaaactc XbaI catalytic APOBEC1 Rv2: acggtctagattattaagggtgccgactcagaaactc), Red color represents the restriction sites. Positive colonies were picked and confirmed by Sanger's Sequencing. The in frame and domain was confirmed using the ExPASY Bioinformatics resource portal and NCBI-BLAST search.

5.2.7 Preparation of the guideRNA:

To prepare the gRNA, a 21 ntd sequence complementary to the target mRNA having a mismatch A at the target C position, was inserted upstream of MS2-RNA by adding the guide sequence with the forward primer of PSL-MS2-6X (pCS2+guide-MS2-RNA) and was ligated with the pCS2+Only vector plasmid under the control of the pol III U6 promoter.

The sequence used for this purpose was as follows: atca**GAATTC***ATTGCTGCGGATCCCATCCAG*GAATGGCCATG. The atca tetra-nucleotide is a leader sequence allowing proper recognition by the restriction enzyme; bold font indicates the restriction site; italic font represents the 21 ntd guide; and the underlined portion indicates the forward primer for MS2-6X. attc**CTCGAG**CGCAAATTTAAAGCGCTGAT (XhoI) is the reverse primer for the

MS2-6X.

Positive colonies were picked and confirmed by Sanger sequencing.

Accordingly, guides with different mismatches were also prepared such as: attcGAATTC**ATTGCTGCGGCTCCCATCCAG**GAATGGCCATG and attcGAATTC**ATTGCTGCGGTTCCCATCCAG**GAATGGCCATG.

Sequence result of the prepared guideRNA:

Green Highlighted parts are the two restriction sites EcoRI and XhoI, **Bold and blue highlighted** part is the guideRNA and the ash color highlighted part is the MS2-6X stem loop.

GAATTC**ATTGCTGCGGATCCCATCCAG**GAATGGCCATGGGACGTCGACCTGAGGTA
ATTATAACCCGGGCCCTATATATGGATCCTAAGGTACCTAATTGCCTAGAAAACATGAG
GATCACCCATGTCTGCAGGTCGACTCTAGAAAACATGAGGATCACCCATGTCTGCAG
TATTCCCGGGTTCATTAGATCCTAAGGTACCTAATTGCCTAGAAAACATGAGGATCAC
CCATGTCTGCAGGTCGACTCTAGAAAACATGAGGATCACCCATGTCTGCAGTATTCC
CGGGTTCATTAGATCCTAAGGTACCTAATTGCCTAGAAAACATGAGGATCACCCATGT
CTGCAGGTCGACTCCAGAAAACATGAGGATCACCCATGTCTGCAGTATTCCCGGGTT
CATTAGATCTGCGCGCGATCGATATCAGCGCTTTAAATTTGCGCTCGAG

5.2.8 Preparation of the Single construct having APOBEC 1 deaminase under the pol II CMV promoter's control and guideRNA under pol III U6 promoter's control:

For the preparation of the single construct the guide RNA construct under the control of the pol III U6 promoter was used as a vector whereas the whole portion of the APOBEC 1 deaminase along with the MS2-HB coat protein part under the control of the pol II CMV promoter up to SV-40 terminator was used as an insert.

The CMV- MS2HB-APOBEC 1-SV40 terminator portion was collected by doing the restriction digestion of the APOBEC1 construct with NdeI and HpaI restriction enzyme.

After the restriction digestion both the vector and insert were made blunt by the Klenow

fragment treatment by using Klenow fragment Kit (Takara). The blunt ended vector and insert were ligated together and after the ligation the direction does not matter as both the parts have their individual promoter and terminator. The final construct was pCS2+CMV+MS2HB+APOBEC1+SV40+U6+guideRNA+MS2stem loop+pCS2-Only. After transformation the positive colonies were extracted by QIAGEN Plasmid Midi Kit (QIAGEN, Germany) and the final construct was sequenced for the final confirmation and the BLAST was done.

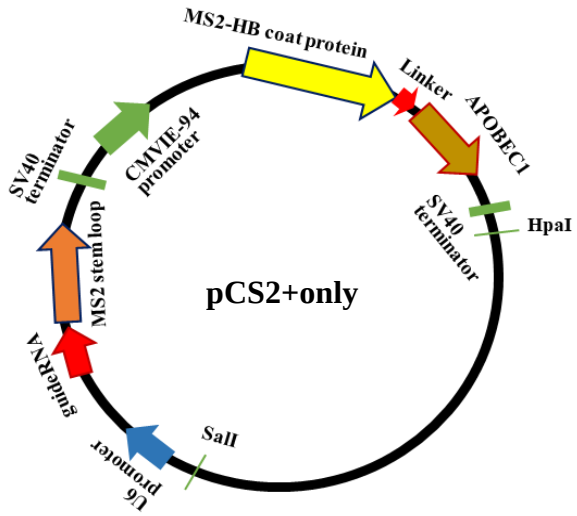


Fig 41. Single construct, the guide portion under the control of the pol III U6 promoter and the APOBEC1 deaminase portion under the control of pol II CMV promoter. Here, CMV promoter is located at the downstream of the U6 promoter

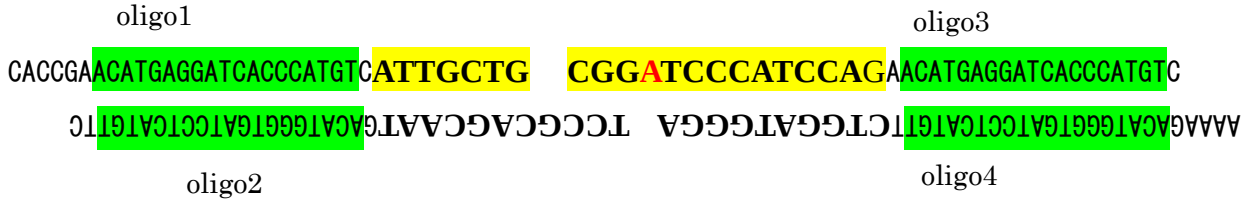
5.2.9 Preparation of 1X MS2 on either side of the guideRNA:

For the preparation of the 1X MS2 on the either side of the guideRNA, four oligos were designed, which was conceptualized by following Katrekar et al., 2019 (14), then the annealing and kination of the oligos was done. Oligo 1+Oligo 2 and Oligo 3+Oligo 4 was annealed and kinated (T4 polynucleotide kinase, 10XC kinase buffer, Biolab). The pol III U6 promoter containing plasmid was digested with the BbsI (New England

Biolabs) restriction enzyme. BAP treatment was done to the digested plasmid. Finally three way ligation was done among the two annealed products and one digested vector. After that the ligated product was transformed into DH5∝ competent cells and positive colonies were collected. Plasmid DNA was extracted by using the Easy pure plasmid extraction kit. For the confirmation of the construct the sequencing of the sample was done.

Oligos	Sequence of the oligos
1	CACCGAACATGAGGATCACCCATGTCATTGCTG
2.	TCCGCAGCAATGACATGGGTGATCCTCATGTTC
3.	CGGATCCCATCCAGAACATGAGGATCACCCATGTC
4.	AAAAGACATGGGTGATCCTCATGTTCTGGATGGGA

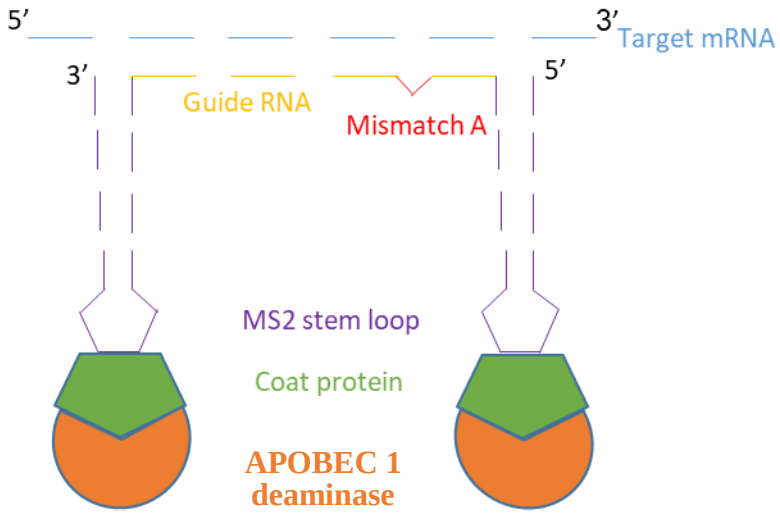
After annealing:



Here yellow highlighted part is the U6 promoter, Green highlighted parts are the two 1XMS2 stem loop on the either side of the guideRNA, pink highlighted part is the guideRNA, rest part is the backbone of pCS2+only plasmid vector.

GGCTGTTAGAGAGATAATTGGAATTAATTTGACTGTAAACACAAAGATATTAGTACA
AAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATG
TTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGCT
TTATATATCTTGTGGAAAGGACGAAACACCGAACATGAGGATCACCCATGTCATTG
CTGCGGATCCCATCCAGAACATGAGGATCACCCATGTCTTTTTTTATGCATGTAATAC
GGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCA
GCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTC
CGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACC
CGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCT

CCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAG
CGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTC
GCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTT
ATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGG
CAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGA
GTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCT
GCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGC
AACA



Katrekar et al., 2019

Fig 42. Schematic model of the 1X MS2 stem loop on either side of the guideRNA mediated editing by APOBEC 1 deaminase

5.2.10 Collection and culturing the Tail fibroblast cells:

ATP7A gene is expressed in different tissues including Fibroblast cells. So I collected the tail fibroblast. For the purpose the 7th Day old Hemizygous male (Ml/y) mice was sacrificed by cervical dislocation. Then immediately the tail of the mice was cut and kept in the ice cold PBS containing petridish for cleaning. Afterwards, a 35 mm dish with DMEM with 1% SP (streptomycin and penicillin) antibiotic and 10% FBS was prepared and the tail part was cleaned and then dipped into the DMEM media followed by gentle chopping by the scissors; so that the cells can easily attach with the bottom of

the dish. After three days the media was changed again with the DMEM with 1% SP (streptomycin and penicillin) antibiotic and 10% FBS. After 7 days when the cells will be properly grown and form a network then the media was subculture of the cells were done and cells were prepared for the transfection. After second or third passages the cells should be seeded for the purpose of electroporation.

5.2.11 Electroporation of the cells for the editing:

Different voltage and conditions were used for the purpose of optimization of the transfection with the Electroporation process. After the electroporation the media was changed after 24 hours with the normal.

The protocol for electroporation for transfecting the gene into the cells:

- The cells were prepared 1 days ago and waited till it became 70% confluent.
- By using Trypsin (Thermofisher) @ 200 ul / well the cells were detached and by pipetting was kept in the solution.
- Then the cell containing solution was taken into the 15 mL tube and more 300 ul of DMEM with 10% FBS (Gibco, USA) was added to the solution, and mixed properly with the solution.
- The solution was centrifuged @ 2000 rpm for 3 mins.
- After the centrifugation the supernatant of the solution was removed
- Then the EP buffer (OptiMEM, Gibco) was added @ 500 ul into each tube, and the precipitated cells were again mixed into the solution of OptiMEM (Gibco)
- Again centrifugation was done @ 2000 rpm for 3 mins
- Then the supernatant was removed
- Again the EP buffer (OptiMEM, Thermofisher) was added @ 200 ul into each tube and mixed well into the solution

- Then 97 ul of the cell containing solution was taken in to the 1.5 mL of Eppendorf tube
- Then total 10 ug of the DNA was added to the solution, total 3 ul
- In total 100 ul (97 ul + 3 ul) of the solution containing the plasmid DNA of the deaminase and the guideRNA was then taken into the cuvette
- Then the cuvette was kept into the electroporation set and then the start button was pressed for the electroporation
- After finishing the electroporation immediately DMEM was added @ 500 ul/ cuvette, then with the supplied dropper the DMEM and electroporated cells were mixed together then in the 24 well plate the cells were seeded along with the DMEM
- After that the cells were incubated into the incubator at 37^0C for 48 hours and then the observation of the cells were done
- Only the control, transfected with the wild type of GFP the expression of the fluorescence was observed because the tail fibroblast neither with the mutation nor without had any fluorescence observation.

Table 7. Application of different pulse rate for electroporation

Poring Pulse	**Transfer pulse**
Voltage: 250 V	Voltage: 20 V
Pulses length: 0.5	Pulses length: 50 msec
Pulse interval: 50 msec	Pulse interval: 50 msec
Pulse no.: 2	Pulse no.: 5

Poring Pulse	**Transfer pulse**
Voltage: 125 V	Voltage: 20 V
Pulses length: 5 msec	Pulses length: 50 msec
Pulse interval: 50 msec	Pulse interval: 50 msec
Pulse no.: 1	Pulse no.: 5

Poring Pulse	**Transfer pulse**
Voltage: 225 V	Voltage: 30 V
Pulses length: 0.5 msec	Pulses length: 50 msec
Pulse interval: 50 msec	Pulse interval: 50 msec
Pulse no.: 2	Pulse no.: 5

Poring Pulse	**Transfer pulse**
Voltage: 275 V	Voltage: 20 V
Pulses length: 0.5 msec	Pulses length: 50 msec
Pulse interval: 50 msec	Pulse interval: 50 msec
Pulse no.: 2	Pulse no.: 5

For transfection with the deaminase and guideRNA

- Total concentration to be transfected through electroporation: 10 ug
- OptiMEM volume: 88 ul
- Deaminase concentration APOBEC1: 262.1 ng/ul
- GuideRNA (CMV with A mis): 2700 ng/ul
- APOBEC 1: 10 ul (2621.0 ng)
- GuideRNA: 2 ul (5400 ng/ul)
- Total DNA concentration in each well: 8021.0 ng/ul

5.2.12 RNA extraction and cDNA synthesis from the electroporated cells:

The transfected dish or plate was first of all was rinsed with ice cold PBS. Then the RNA was collected by the TRIzol® reagent (Invitrogent) method followed by cDNA was synthesis with Superscript ™ III First stand synthesis system (Invitrogen) according to the protocol given by the manufacturer.

5.2.13 PCR amplification of the synthesized cDNA and Sanger's sequencing:

For the PCR from the synthesized cDNA Fw primer: CTGGATGTTGTGGCAAGTATTGAC; Reve primer: GCTGTTCAGGGAGCGCTTG were used. The total length of the amplified fragment was 466 bp.

5.2.14 Sequence confirmation of the editing:

After the PCR amplification, the amplified DNA (PCR product) was run in 6% polyacrylamide gel by loading equal volume (3 ul) of PCR product into each well. The gel was observed in LAS 3000 gel imager. Finally the observed band was cut and DNA was purified by using the Gel purification kit. Then the purified sample was sent for the Sanger's sequencing. The sequence data was analyzed and editing rate was calculated from the peak height and peak area.

Editing efficiency (sense) =

$$\text{Considering peak Area: } \frac{\text{Area of T}}{\text{Area of T + Area of C}} \times 100$$

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

5.3 RESULTS

5.3.1 Body weight difference between the hemizygous male, heterozygous female and normal male littermate at 7, 10 and 14 days:

The weight was measured at three different days of age such as: 7, 10 and 14 days. I found that at the age of 10 days the weight of all the mice: normal littermate, heterozygous female and also the hemizygous male were increased comparing with the body weight of 7 days, p value was 0.0039, which is less than 0.05, so the weight increased significantly from Day 7 to Day 10 (Figure 43).

Onwards, on Day 14, for the heterozygous female and normal littermate weight increased but in case of the hemizygous male weight started to decrease. The p value was 0.002, which is lesser than 0.01. So the decreased value of the weight is significant at 95% confidence interval.

Koyama et al., 1991 in their paper presented that no significant difference in body weight was found through days 7-14 in macular males (Ml/y). Although the body weight of normal littermates (+/y) at 7 days of age was almost same as that of the Ml/y, the former gained more weight than the layer at 10 days of age, and no significant increase was observed after 10 days of age. The statement of the paper supported my observation of decreasing weight after 10 days (Figure 43).

Table 8. Weight measurement of the macular mouse at different ages

Heterozygous female (Ml/+)

Normal males (+/y)

Hemizygous male mice (Ml/y)

Days	Genotype	Body weight (g)	Average+/-STDEV
7	(Ml/+)	4.2, 4.3, 4.5	4.33±0.15
	(+/y)	4.3, 4.4, 4.6	4.43±0.12
	(Ml/y)	3.8, 3.7, 3.6	3.7±0.08
10	(Ml/+)	6.2, 6.1, 6.5	6.2±0.17
	(+/y)	6.4, 6.3, 6.5	6.4±0.082
	(Ml/y)	4.2, 4.1, 4.6	4.3±0.27*
14	(Ml/+)	6.8, 6.7, 6.6	6.7±0.1
	(+/y)	7.3, 7.0, 7.5	7.27±0.25
	(Ml/y)	3.4, 3.2, 3.1	3.2±0.2**

*=the p value is significant at 95% confidence interval, as p< 0.05
**= the p value is significant at 95% confidence interval as p<0.01

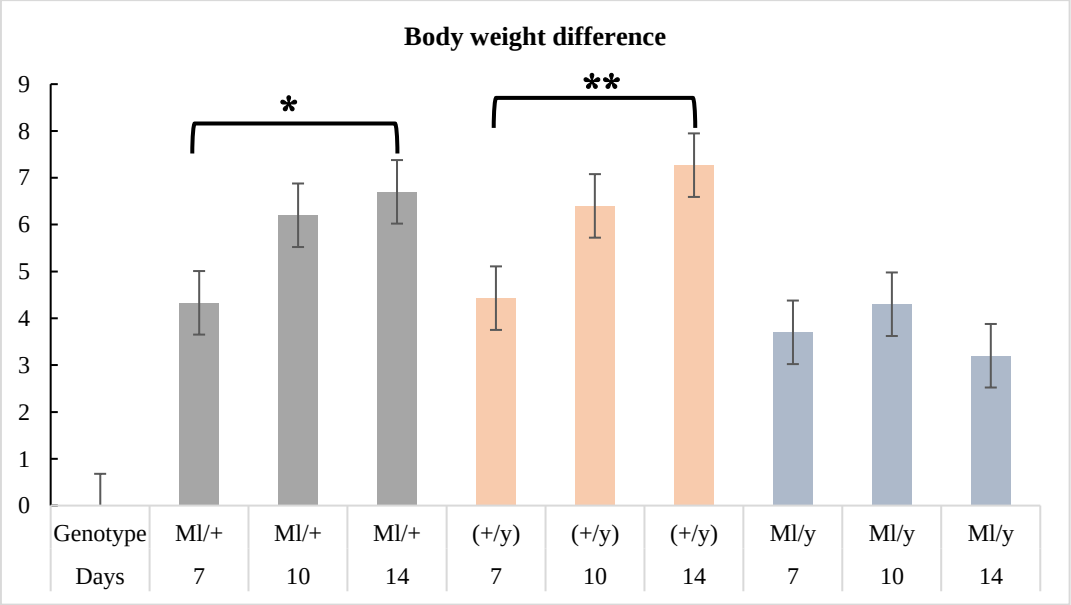


Fig 43. Body weight change along with days according to the genotype. The graph shows that with the progress of the days the body weight of the heterozygous female (Ml/+) and normal littermate male (+/y) increased. But in case of the hemizygous male (Ml/y), which is the macular mouse, the body weight increased upto 10 days but after that at 14th days it started to reduce which became even less than the normal bodyweight at 7th day. For all data the statistical analysis (mean±SEM) was done, where n=3.

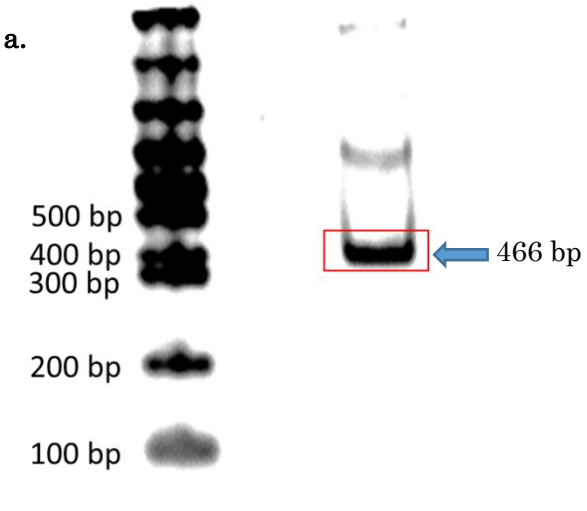
5.3.2 Sanger's sequencing confirmation of the genetic code restoration:

As there was no fluorescence expressing gene into the cells where the transfection was done neither by any fluorescence containing plasmid was transfected. Moreover, there was no restriction site as well. As a result the evidence of the restoration was completely dependent on the Sanger's sequencing.

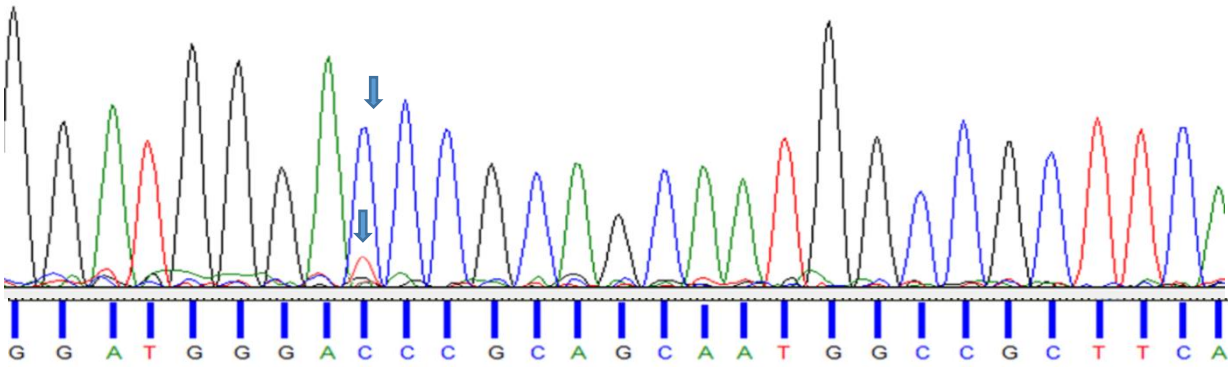
The Sanger's sequencing analysis gave the evidence of the restoration of the genetic code. As in case of the experimental sample at the position of the target C a small peak of restored T was also found, which provided the evidence of the restoration of genetic code. In case of the mutated sample only a single peak of C was observed but after the transfection with the APOBEC1 deaminase and guideRNA, editing happened as a result the dual peak of mutated C and restored T was observed which means that the deaminase system worked for the restoration of the genetic code for T to C mutation. Whereas when only one factor was transfected either the APOBEC 1 deaminase or the guideRNA, no such restoration was found as only the mutated C peak was observed no restored T peak was there (Figure 44).

From the peak area and also the peak height the editing efficiency was calculated. The peak area and peak height from the Sanger's sequencing analysis was measured by ImageJ software (NIH). From the calculation it was found that by using the APOBEC 1 deaminase and U6-21bp upstream-MS2-6X 12.1% and 16.25% of the genetic code was restored by peak area and peak height, respectively (Figure 44).

a.



b.



c.

	Peak height	Peak width	Peak area	Editing efficiency	
Mutated C	134	21	2814	T/(T+C)*100	
Restored T	26	15	390	0.121723	0.1625
				12.17%	16.25%

d.

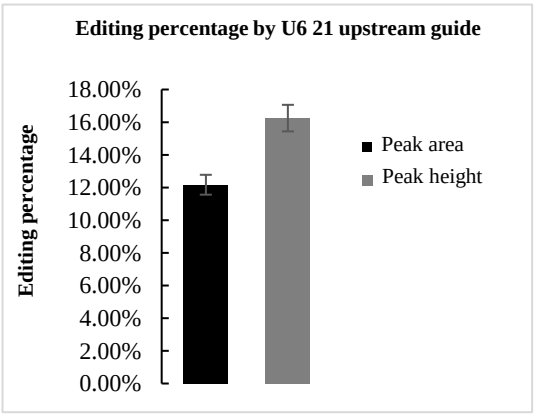


Fig 44. Confirmation of the restoration of the genetic code where a. PCR amplification of the ATP7A gene from the synthesized cDNA of transfected cells b. Sanger's sequencing showing the restoration of the genetic code. At the position of the mutated C there is a small peak of T (Arrow), which is due to the RNA editing, c. calculation of the editing efficiency from the peak area and peak height and d. graph showing that according to peak area and peak height 12.17 % and 16.25% has been restored, respectively. For all data the statistical analysis (mean±SEM) was done, where n=3.

Later on, the single construct made of the U6 promoter containing guideRNA construct was applied, where the guide part was placed upstream of the MS2 stem loop and CMV APOBEC 1 deaminase in a single plasmid vector. In this case the CMV promoter was placed upstream of the U6 promoter. Then the single construct was transfected into the fibroblast cells of the macular mouse of hemizygous male by the electroporation method.

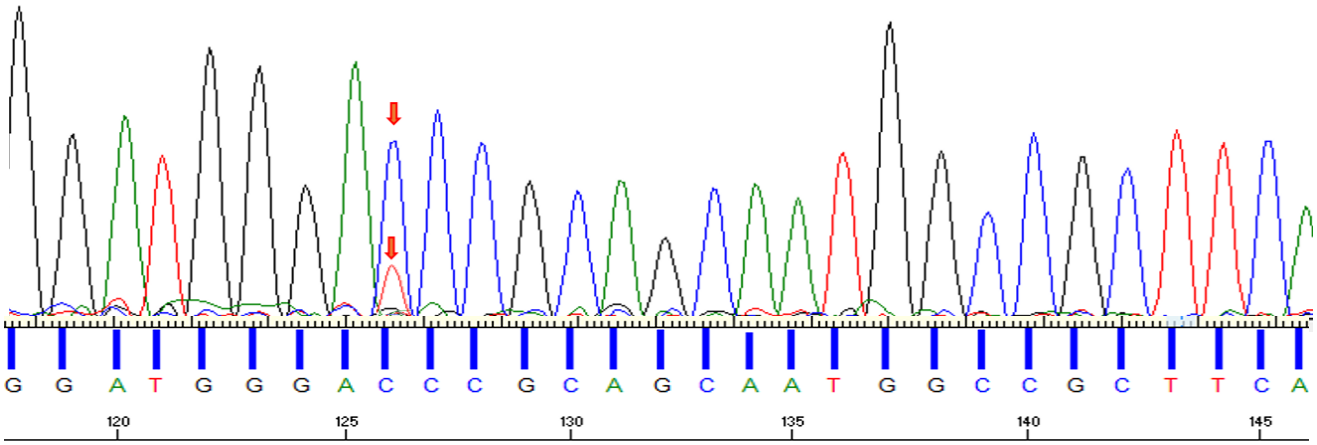
After the transfection, in the similar way the RNA was extracted from the cells and followed by the cDNA synthesis was done. Later on the PCR amplified fragment was sequenced with the sense primer for the Sanger's sequencing analysis.

In this case also the Sanger's sequencing analysis gave the evidence of the restoration of the genetic code. In case of the experimental sample at the position of the target C a small peak of restored T was also found, which evident the restoration of the genetic code. In case of mutated sample only a single peak of C was observed but after the transfection with the single construct having the APOBEC1 deaminase and guideRNA, the editing happened as a result the dual peak of mutated C and restored T was observed which means that the deaminase system worked for the restoration of the genetic code.

From the peak area and also the peak height I calculated the editing rate. The peak area and peak height from the Sanger's sequencing analysis was measured by ImageJ software.

From the calculation by peak area and peak height I found that by using the APOBEC 1 deaminase and U6-MS2-6X 21 upstream, 27.20% and 26.09% of the genetic code was restored, respectively (Figure 45).

a.



b.

	Peak height	Peak Width	Peak area	Editing percentage	
Mutated C	119	17	2023	T/(C+T)*100	
Edited T	42	18	756	27.20	Peak area
				26.09	Peak height

c.

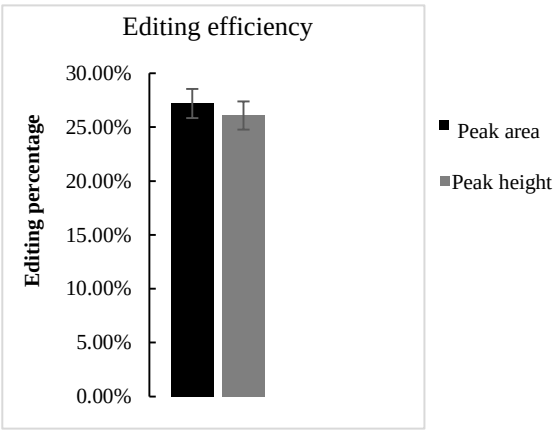


Fig 45. Confirmation of the restoration of the genetic code where by using single construct a. Sanger's sequencing showing the restoration of the genetic code. At the position of the mutated C there is a small peak of T (Arrow), which is due to the RNA editing, b. calculation of the editing efficiency from the peak area and peak height and c. graph showing that according to peak area and peak height 27.20 % and 26.09% has been restored, respectively. For all data the statistical analysis (mean±SEM) was done, where n=3.

5.3.3 Application of GuideRNA with 1X MS2 on either side of the guideRNA (Double MS2):

Later on the 1X MS2 containing guideRNA construct was used for the purpose of

editing along with the APOBEC1 deaminase in the macular mouse tail derived fibroblast cells using the electroporation method. For the purpose of electroporation of the editing factors several different pulses were used for the optimization of the condition. They are as follows:

Poring Pulse		**Transfer pulse**
Voltage	120 V	15 V
Pulse length	0.5 msec	50 msec
Pulse Interval	50 msec	50 msec
Pulse Number	2	2

Poring Pulse		**Transfer pulse**
Voltage	150 V	20 V
Pulse length	1 msec	50 msec
Pulse Interval	50 msec	50 msec
Pulse Number	2	5

Poring Pulse		**Transfer pulse**
Voltage	250 V	20
Pulse length	1 msec	50 msec
Pulse Interval	50 msec	50 msec
Pulse Number	2	5

Poring Pulse		**Transfer pulse**
Voltage	150 V	15 V
Pulse length	0.5 msec	50 msec
Pulse Interval	50 msec	50 msec
Pulse Number	2	5

Poring Pulse		**Transfer pulse**
Voltage	120 V	20
Pulse length	0.5 msec	50 msec
Pulse Interval	50 msec	50 msec
Pulse Number	2	5

After the optimization of the condition I have found that for this type of guideRNA 150 V is good enough as the cell condition was better in this case, as a result I choose this condition of electroporation.

After 48 hours of electroporation and incubation, the RNA from the cells were extracted followed by cDNA synthesis. Then PCR amplification was done and the amplified DNA was run in PAGE, afterwards the band was cut and purified DNA was sequenced for the confirmation of the editing.

From the sequence data I could see the small restored T peak at the position of the mutated C. Editing rate was calculated from the peak area and peak height, which was 36.66% and 34%, respectively (Figure 46).

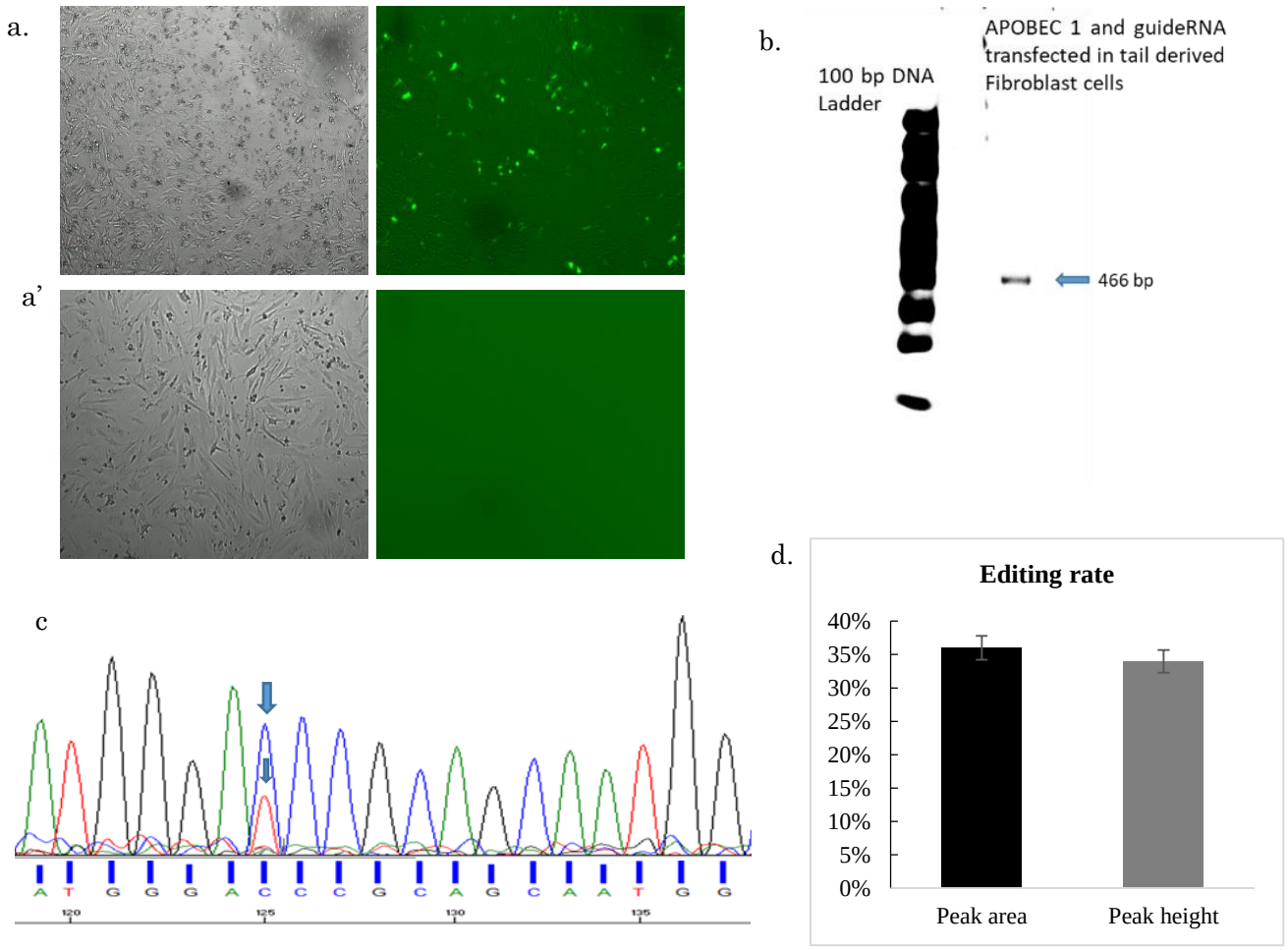


Fig 46. a. Fluorescence image showing the transfection efficiency while the wild type GFP was transfected into macular mouse tail derived fibroblast cells having mutated ATP7A gene, a'. on the contrary when the APOBEC1 and guideRNA was transfected into the macular mouse tail derived fibroblast cells having mutated ATP7A gene there was no fluorescence expression as there was no fluorescence expressing gene, image was taken by Juli fluorescence microscope b. PCR amplification of the targeted ATP7A gene from the predicted restored sample, transfected with APOBEC 1 deaminase and guideRNA c. Sanger's sequencing confirmation of the C to U editing, by using the 1X MS2 guideRNA along with APOBEC 1 deaminase in macular mouse tail derived Fibroblast cells. a. Sequence data showing restored T peak at the position of mutated C peak. d. calculation of the editing rate from the peak area and peak height, respectively. For all data the statistical analysis (mean±SEM) was done, where n=3.

5.4 DISCUSSION

5.4.1 Body weight reduction in case of the Macular Mouse:

I have found from the data that all the heterozygous female (Ml/+), normal littermate male (+/y) and hemizygous male (Ml/y) have increased the weight as usual up to 10 days. After that the body weight of heterozygous female (Ml/+), normal littermate increased significantly at 14 days as well with a p value of 0.0039, but in case of the hemizygous male (Ml/y) the body weight decreased significantly with a p value of 0.002. It became even less than the body weight at 7^{th} day. This feature is identical with the findings of Koyama et al., 1991, who have found that the body weight of normal littermates (+/y), and heterozygous female (Ml/+) at 7 days of age was almost the same as that of macular mouse (Ml/y), the former gained more weight than the latter at 10 days of age, and no significant increase was observed after 10 days. Yamano et al., 1987 (13) also observed that hemizygotes (Ml/y) increased at the same rate as the others' until day 10, but thereafter, showed gradual weight loss.

5.4.2 Sanger's sequence confirmation of the RNA editing:

The Sanger's sequencing analysis gave the evidence of the restoration of the genetic code. As in case of the experimental sample at the position of the target C a small peak of restored T was also found, which provides the evidence of the restoration of the genetic code. In case of the only mutated sample only a single peak of C was observed but after the transfection with the APOBEC1 deaminase and guideRNA, editing happened as a result the dual peak of mutated C and restored T was observed which means that the deaminase system worked for the restoration of the genetic code for T to C mutation. Whereas when only one factor was transfected either the APOBEC1 deaminase or the guideRNA, no such restoration was found as only the mutated C peak

was observed no restored T peak was there.

From the peak area and also the peak height the editing efficiency was calculated. The peak area and peak height from the Sanger's sequencing analysis was measured by ImageJ software (NIH). From the calculation I found that by using the APOBEC 1 deaminase and U6-21bp upstream-MS2-6X 12.17% and 16.25% of the genetic code was restored, respectively by peak area and peak height.

In case of the single construct transfection also the Sanger's sequencing analysis gave the evidence of the restoration of the genetic code. As in case of the experimental sample at the position of the target C a small peak of restored T was also found, which evident the restoration of the genetic code. In case of the only mutated sample only a single peak of C was observed but after the transfection with the single construct having the APOBEC1 deaminase and guideRNA, the editing happened as a result the dual peak of mutated C and restored T was observed which means that the deaminase system worked for the restoration of the genetic code.

From the peak area and also the peak height I calculated the editing rate. The peak area and peak height from the Sanger's sequencing analysis was measured by ImageJ software.

From the calculation I found that by using the APOBEC 1 deaminase and U6-MS2-6X 21 upstream 27.20% and 26.09% of the genetic code was restored, respectively by peak area and peak height.

Regarding the effectiveness of the single construct with a combination of the CMV and U6 promoters, the experimental data supported the findings of Jianguo Su et al., (2008) (22), who found that when the hybrid construct produced where the CMV promoter or enhancer is placed to the immediate upstream of the U6 promoter, the silencing

efficiency by the shRNA increased. The hybrid construct of CMV deaminase and U6 guideRNA is also prepared where the CMV was placed immediately to the upstream of the U6 promoter. And, according to the findings of Jianguo Su et al., group I have also found an improved editing efficiency comparing with the individual construct of APOBEC 1 and guideRNA constructs in two different plasmids.

The probable reason behind it could be, as in case of the individual construct of deaminase and guideRNA, both the factors are applied individually where the ratio may not be maintained properly and also two different factors when transfected they will create much more stress to the cells than a single construct transfection. Furthermore in case of the single construct there is no need to maintain any ratio between the deaminase and gudieRNA, as both of them are together in the same construct, so it's easy to do the transfection by giving less stress to the cells. On the other hand a combination of pol III U6 promoter and pol II CMV promoter or enhancer elevates the editing efficiency comparing to a single promoter either pol II CMV or pol III U6 alone.

Afterwards, the 1XMS2 containing guideRNA was introduced along with the APOEBC1 deaminase. Similarly the sample was sequenced for observing the editing rate. Editing rate was calculated both by peak area and peak height. I have found that editing rate was 36.66% and 34%, respectively, which was also supported by the study of Katrekar et al., (2019). They found that the Double MS2 system works better than other stem loop systems and addition of a NES with the construct would have increased the editing rate much more (24). So in the near future that can be applied for the purpose.

5.5 CONCLUSION

This investigation would be critical to help orderly improvement in the specificity and safety of this approach. Another significant thought while considering RNA targeting for gene therapy, especially via the use of non-integrating vectors, is the need for intermittent re-administration of the effector constructs and the other transcripts owing to the typically limited half-life of edited mRNAs and effectors, which could be easily done by using the viral vectors or liposome or lipid mediated delivery system. In this regard, compared to CRISPR-based RNA editing approaches, the RNA-guided deaminase strategy is directly relevant to human therapeutics because versions of this toolset solely utilize effector RNAs and human proteins. Additionally, as deaminases are widely expressed for instance, ADAR1 across most human tissues and ADAR2 in particular in the lung and brain, APOBEC1 only in small intestine in case of human-endogenous recruitment of these via adRNAs bearing long-antisense domains presents a very attractive strategy for efficacious RNA editing. I thus anticipate that with progressive improvements, this tool set would have broad implications for diverse basic science and therapeutic applications. In case of the several other effector systems such as CRISPR-Cas system, are of prokaryotes origin, raising a significant risk of immunogenicity for *in vivo* therapeutic applications. To avoid these limitations of DNA nucleases, approaches that instead directly target RNA would be highly desirable, as these would enable tenability, reversibility and most importantly no off-target mutations would be permanent. Additionally, RNA unlike DNA, can be targeted via simple RNA-nucleic acid hybridization. Thus, RNA base editing via RNA guided cytidine deaminases of the human origin could be an attractive approach for *in vivo* correction of disease causing point mutations.

5.6 REFERENCES

1. Das S., Levinson B., Vulpe C., Whitney S., Gitschier J., & Packman S. Similar Splicing Mutations of the Menkes/Mottled Copper-Transporting ATPase Gene in Occipital Horn Syndrome and the Blotchy Mouse. Am. J. Hum. Genet. 56: 570-576, (1995).
2. Mori M., & Nishimura M. A serine-to-proline mutation in the copper-transporting P-type ATPase gene of the macular mouse. Mammalian Genome 8:407-410, (1997).
3. Menkes J.H.M., Alter M., Steigleder G.K., Weakley D.R., & Sung J.H. A sex-linked recessive disorder with retardation of growth, pericular hair and focal cerebellar degeneration. Pediatrics 29: 764-779, (1962).
4. Lutsenko S., Washington-Hughes C., Ralle M., & Schmidt K. Copper and the brain noradrenergic system. J Biol Inorg Chem. 24(8):1179-1188, (2019).
5. Kaler S.G. Menkes disease. Adv Pediatr 41: 263-304, (1994)
6. Chelly J., Turner Z., Tonnesen T., Petterson A., Ishikawa-Brush Y., Tommerup N., Horn N., & Monaco A.P. Isolation of a candidate gene for Menkes disease that encodes a potential heavy metal binding protein. Nature Genet 3: 14-19 (1993).
7. Mercer J.F., Livingston J., Hall B., Paynter J.A., Begy C., Chandrasekharappa S., Lockhart P., Grimes A., Bhave M., Siemieniak D., & Glover T.W. Isolation of a partial candidate gene for Menkes disease by positional cloning. Nature Genet 3: 20-25, (1993).
8. Vulpe C., Levinson B., Whitney S., Packman S., & Gitschier J. Isolation of a candidate gene for Menkes disease and evidence that it encodes a copper-transporting ATPase. Nature Genet 3: 7-13, (1993).
9. Mercer J.F., Grimes A., & Ambrosini L. Mutation sin the murine homologue of the

Menkes gene in dappled and blotchy mice. Nature Genet 6: 374—378, (1994).

10. Levinson B., Vulpe C., Elder B., Martin C., & Verley F. The mottled gene is the mouse homologue of the Menkes disease gene. Nature Genet 6: 369-373, (1994).
11. Barnes N., Tsivkovskii R., Tsivkovskaia N., & Lutsenko S. The copper-transporting ATPases, menkes and wilson disease proteins, have distinct roles in adult and developing cerebellum. J Biol Chem. 280 (10): 9640-5, (2005).
12. Ohta Y., Shiraishi N., & Nishikimi M. Occurrence of two missense mutations in Cu-ATPase of the macular mouse, a Menkes disease model. Biochem Mol Biol Int; 43(4): 913-8, (1997).
13. Yamano T., Shimada M., Kawasaki H., Onaga A., & Nishimura M. Clinico-pathological study on macular mutant mouse. Acta Neuropathol (Bed), 72: 256-260, (1987).
14. Koyama M., Ishida T., Horiike K., Nozaki M., & Shimada M. Urate Oxidase Activity and Copper Content in the Liver of Macular Mutant Mouse, a Model Animal for Human Congenital Copper Deficiency, Menkes' Kinky Hair Disease. J. Nutr. Sci. Vitaminol, 37: 601-609, (1991).
15. Packman S., Palmiter R.D., Karin M., and O'Toole C. Metallothionein messenger RNA regulation in the mottled mouse and Menkes kinky hair syndrome. J Clin Invest. 79(5): 1338–1342, (1987).
16. Katsura T., Kawasaki H., Yamano T., and Himada M. Copper contents and pathological changes in various organs of macular mouse. Congenital Anom., 28: 85-92, (1988).
17. Kumode M., Yamano T., Shimada M. Neuropathologicai study on cerebellum of macular mutant mouse heterozygote. Acta Neuropathol, 86:411-417, (1993).

18. Katrekar D., Chen G., Meluzzi D., Ganesh A., Worlikar A., Shih Yu-Ru., Varghese S., and Mali P. In vivo RNA editing of point mutations via RNA guided adenosine deaminases. Nature Methods, 16(3): 239–242, (2019).
19. Seki K., Sato T., Ishigaki Y., Nakamura S., Ishihara Y., and Ozawa T. Decreased activity of cytochrome c oxidase in the macular mottled mouse: an immuno-electron microscopic study. Acta Neuropathol, 77: 465-471, (1989).
20. Suzuki–Kurasaki M., Okabe M., and Kurasaki M. Copper–metallothionein in the Kidney of Macular Mice: A Model for Menkes Disease. The Journal of Histochemistry & Cytochemistry, 45(11): 1493–1501, (1997).
21. Skjorringe T., Pedersen P.A., Thorborg S.S., Nissen P., Gourdon P. & Moller L.B. Characterization of ATP7A missense mutants suggests a correlation between intracellular trafficking and severity of Menkes disease. Scientific Reports, 7: 757, (2017).
22. Donsante A., Yi L., Zerfas P.M, Brinster L.R., Sullivan P., Goldstein D.S., Prohaska J., Centeno J.A., Rushing E. and Kaler S.G. ATP7A Gene Addition to the Choroid Plexus Results in Long-term Rescue of the Lethal Copper Transport Defect in a Menkes Disease Mouse Model. Molecular Therapy 19(12): 2114–2123, (2011).
23. Iwase T., Nishimura M., Sigimura H., Igarashi H., Ozawa F., Shinmura K., et al.. Localization of Menkes gene expression in the mouse brain; its association with neurological manifestation in Menkes model mice. Acta Neuropathol. 91: 482–488, (1996).

CHAPTER: VI

Final Discussions

At present, genome editing has become very popular and obtained much attention after the establishment of the CRISPR-Cas system although other programmable genome editing tools like Transcription activator-like effector nucleases (TALEN), Zinc Finger Nuclease (ZFN) reporter are still being used mainly for developing different genetically modified organisms (GMO) for the purpose of research. Engineering of different gene editing tools largely depend on calculations of designing guide RNA sequence, avoiding off-target effects and most importantly targeting specificity calculation etc. Editing efficiency, specificity and safety are of big deals in case of the utilization of genome editing tools. For CRISPR-Cas system homology-mediated repair system is needed to repair between cut and inserted DNA which is only present in dividing cells. In case of the non-dividing cells like nerve and muscle cells, this system facing challenges. In association with the technical challenges and bioethical issues (1), it is still controversial regarding the clinical application of this technology. Because in case of the genetic diseases only one gene or part of the gene is predicted to be defective but when CRISPR system is being applied into the germ cells or embryo, other genes may also have the possibilities to be mutated thus ultimately it may affect multi organ functionality. CRISPR system in association with the deaminase is also used for the single nucleotide substitution at DNA level (2). However, since its introduction in 2012, CRISPR gene editing has held the guarantee of relieving over 6,000 known genetic diseases. Presently it's being scrutinized. Both China and USA has made the genuine applications of the CRISPR on the human for the purpose of treatment. In the initial spate of clinical trials, scientists are using CRISPR/Cas9 to combat cancer and blood disorders in people. In these researches, researchers expel some of a person’s cells, edit the DNA and then inject the cells back in, now hopefully armed to combat disease. Researchers are also set

to observe how CRISPR/Cas9 acts inside the human body. In an upcoming trial, individuals with an inherited defect in vision will have the molecular scissors infused into their eyes. But big questions remaining about whether CRISPR/Cas9 can satisfy everyone's expectations or not, because other previously promising technologies have missed the mark. For instance, stem cell injections helped deadened rats to walk once again. But they didn't act up to the mark for people. Although conventional gene therapies, which insert healthy copies of genes to supplant or counteract disease-causing variants, likewise endured extreme setbacks. Irrespective of all the potential risks that could be occurred through CRISPR, still, CRISPR is more precise than traditional gene therapy and therefore may have the potential to treat some diseases for which gene therapy hasn't worked admirably. CRISPR's notoriety was defamed last year after a researcher in China edited a gene in embryos that went on to development of two baby girls in 2018, which caused a provocation to the ethical challenges for using CRISPR. The current CRISPR trials don't have the same ethical issues - the therapies are being experimented in grownups and children, and won't prompt the DNA changes that can be inherited.

At long last, several effector techniques, such as the CRISPR-Cas systems, are derived from prokaryotic origin, raising a noteworthy risk of immunogenicity for *in vivo* remedial application. To avoid these confinements of DNA nucleases, moves toward that rather straightforwardly target RNA would be exceptionally preferable, as these would enable tenability and reversibility and critically no off-target effects would become permanent. Moreover, RNA, unlike DNA can be targeted through simple RNA nucleic acid hybridization. In this manner, RNA base editing be means of RNA guided adenosine or cytidine deaminases of human origin could be an alluring mechanism for

in vivo correction of the point mutations which may cause diseases. RNA editing is one of the most significant post-transcriptional alterations of the genetic code information encoded in the genome of a living organism. There are distinctive group of deaminases those are responsible for the A-to-I, C-to-U and even U-to-C type of RNA editing. Artificial RNA editing is the process where only the targeted mRNA is being edited for the aim of tuning protein action without destroying the highly composed and complexed genomics DNA. Naturally, the deaminase ties to the target by sequence explicit manner with the aid of its related structures. For example, ADARs bind to the dsRNA framed by the altered Alu repeats in RNA whereas APOBEC target the cytidine (C) adjacent to the Mooring sequence. The deaminases are programmable to the target specific nucleotide at any sequence. Harnessing and engineering the natural deaminase is an excellent prospect for the therapeutic application due to the purpose of treating patients having the point mutated genetic diseases in future.

In the **Chapter II**, It showed that the recognized MS2 system can be utilized to control the deaminase domain of the ADAR1 towards the site specific A-to-I RNA editing (4-5). The chimeric protein of MS2-ADAR1-DD and genetically encoded guideRNA express well in HEK 293 cells, which become active after being expressed. Fluorescent microscopic results indicated that the signals in experimental wells were resembling with the wild type EGFP which proved that this developed deaminase system is capable of converting the mutated Ochre stop codon (TAA, G-to-A mutation) to restored GFP (TGG). Restriction Fragment Length Polymorphism (RFLP) analysis clearly indicated that the RNA editing happened exactly at the targeted position. Sanger's sequencing of the target substrate proven as a good sort of evidence regarding the specificity of the editing by dual peak at the targeted adenosine (A) site. The Sanger's sequencing was

also performed with both the sense and antisense primers, and the editing rate was calculated from the peak area and peak height, measured by the Image J software. Only the positive cells were selected and collected as a result around 93% of the 5'A and 86% of the 3'A of the mutated TAA (double mutation is there so double restoration is needed) was restored to TGG. Later on the Double-MS2 guideRNA (1X-MS2 stem loop containing on the either side of the guideRNA) (15) applied on the Ochre stoop codon (TAA) along with the ADAR1 deaminase domain, where we obtained around 26% editing efficiency in case of the 5'A and around 18% of editing efficiency in case of 3'A. In this case, whole dish of cells were collected including both the positive means expressing and negative means non-expressing cells.

In **Chapter III**, an artificial RNA editase system was engineered by combining the deaminase domain of APOBEC1 (apolipoprotein B mRNA editing catalytic polypeptide 1) with a guideRNA (gRNA) which is complementary to the target mRNA sequence. In this artificial enzyme system, gRNA was bound to the MS2 stem-loop, and deaminase domain, which has the ability to convert mutated target nucleotide from C-to-U, was fused to MS2 coat protein. As a target RNA, the RNA encoding blue fluorescent protein (BFP) was used, which was derived from the gene encoding GFP by 199T > C mutation. Upon transient expression of both components (deaminase and gRNA), the GFP fluorescence expression was observed by LSM confocal microscopy, indicating that mutated 199th C in BFP had been converted to U, restoring original sequence of GFP. This result was confirmed by PCR-RFLP and Sanger's sequencing using cDNA from transfected cells, revealing an editing efficiency of approximately 21%. Deep RNA sequencing result showed that off-target editing was negligible in this system. Successful development of an artificial RNA editing system using artificial deaminase

(APOBEC1) in combination with MS2 system could lead to therapies that treat genetic diseases by restoring the mutated sequence to wild-type sequence, at the mRNA level.

In the **Chapter IV**, I tried to make a difference between the pol II and pol III promoters and also the single construct in respect to editing efficiency from C-to-U restoration (6-10). From the experimental data it was found that in case of the CMV promoter controlled process of RNA editing where both the deaminase and guideRNA constructs were prepared under the control of the pol II CMV promoter, the editing efficiency was lesser comparing to the U6 promoter containing guideRNA or in single construct having combined approach of CMV with deaminase and U6 promoter with guideRNA construct in a single plasmid vector. It was also found from the PCR-RFLP (band intensity) data that with the increased concentration of the deaminase or the guideRNA, the restoration percentage also increased. But the main concern was the off-target effect. Because from one of the previous experiments of our group it was found that if the concentration of the deaminase domain was increased per well during transfection then there was a possibility that the editing rate may increase but at the same time it may increase the off-target effects as well. So in the future work I will consider this particular issue of off-target effect which I haven't done yet. I also tried to calculate the editing efficiency from the peak area of the Sanger's sequencing data. After the calculation of the efficiency it was found that in case of the CMV controlled approach the rate was 21.02% whereas in case of the U6 controlled and single construct the restoration rate was 39.37% and 41.65%, respectively. If I compare the editing efficiency rates achieved between the CMV controlled guideRNA approach and U6 controlled guideRNA approach along with the CMV controlled APOBEC 1 deaminase the p Value is 0.0044, in case of U6 controlled guide approach the editing efficiency

significantly increased than that of CMV approach. Again regarding the comparison between Single construct and the U6 promoter p value was 0.0007 which was highly significant, so in case of the single construct the editing efficiency increased significantly (11-14).

In **Chapter V**, I have tried to apply the system in the real mouse model. The macular mouse is a mutant strain that emerged in C3Hf stock and has a different mutant allele (the old symbol was Mo^{ml}, and the new symbol is ATP7A $Mo^{\sim ml}$) on the mottled locus. The macular mouse is considered to be an ideal model for human Menkes disease. The macular mouse is contemplated to be an ideal model for human Menkes disease (16).

I have found from the data that all the heterozygous female (Ml/+), normal littermate male (+/y) and hemizygous male (Ml/y) have increased the weight as usual up to 10 days. After that the body weight of heterozygous female (Ml/+), normal littermate increased significantly at 14 days as well with a p value of 0.0039, but in case of the hemizygous male (Ml/y) the body weight decreased significantly with a p value of 0.002. It became even less than the body weight at 7^{th} day (16, 17). The peak area and peak height from the Sanger's sequencing analysis was measured by ImageJ software. From the calculation I found that by using the APOBEC 1 deaminase and U6-21bp upstream-MS2-6X 12.17% and 16.25% of the genetic code was restored, respectively by peak area and peak height. From the peak area and also the peak height I calculated the editing rate. The peak area and peak height from the Sanger's sequencing analysis was measured by ImageJ software. From the calculation I found that by using the APOBEC 1 deaminase and U6 MS2-6X-21bp upstream 27.20% and 26.09% of the genetic code was restored, respectively by peak area and peak height.

Afterwards, the 1XMS2 containing guideRNA (15) was introduced along with the

APOBEC1 deaminase. Similarly the sample was sequenced for observing the editing rate. Editing rate was calculate both by peak area and peak height. I have found that editing rate was 36% and 34%, respectively.

In conclusion, in my developed artificial RNA editing system in the fibroblast cells derived from the tail of the macular mouse I have found the highest 36% of editing efficiency after application of the double MS2-1X stem loop containing guideRNA under the control of pol III U6 promoter along with the APOBEC1 deaminase under the control of the pol II CMV promoter. However the editing efficiency and specificity are the most important and biggest challenges, for the sake of the matters many factors needed to be considered. Although the work is still under progress. Where I am trying with the different types of guideRNA considering variation in length and complementary sequence as well as different number of stem loops to be used.

Another thing, the MS2-deaminase containing plasmid pCS2+MT-MS2HB-APOBEC-DD has Myc-tag and HB tag peptides, where Myc-tag is located upstream of this chimeric protein and HB located between the MS2-deaminase. HB tag is approximately 300 bp in length. Therefore these structures might affect the efficiency of the whole system. So in the near future I can cut off that His-Bind (HB) tag part and can evaluate the activity of this system via editing efficiency and also specificity.

Although for any developed system delivery to the patient is very significant. To date AAVs have been the most appropriate for this purpose because of their capacity to infect a variety of cells, which in turn allows efficient delivery to a broader range of tissues, especially the liver, muscle, eye and heart. Furthermore, although long term expression of the RNA editing parts might be required to lengthen the therapeutic effects, at the

same time cytotoxicity should also be checked. So in the near future I will apply the developed system into the real model mice for checking the efficacy *in vivo* by using the (AAV) viral vector delivery technique. The successful application of the developed artificial enzymatic system will open a new window in this realm.

REFERENCES

1. Kim D., Bae S., Park J, Kim E, Kim S, Yu HR et al. Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. Nat Methods, 12: 237–243, (2015).
2. Gaudelli NM., Komor AC., Rees HA., Packer MS., Badran AH., Bryson DI. Programmable base editing of A-T to G-C in genomic DNA without DNA cleavage. Nature, 551: 464–471, (2017).
3. Vogel P., and Stafforst T. Site-directed RNA editing with antagomir deaminases-A tool to study protein and RNA function. ChemMedChem, 9: 2021–2025, (2014).
4. 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).
5. Paul C.P. Effective expression of small interfering RNA in human cells. Nat. Biotechnol. 20: 505–508, (2002).
6. Xia X.G. An enhanced U6 promoter for synthesis of short hairpin RNA. Nucleic Acids Res. 31:e100, (2003).
7. Ong S.T. Hybrid cytomegalovirus enhancer-h1 promoter-based plasmid and baculovirus vectors mediate effective RNA interference. Hum. Gene. Ther. 16:1404–1412, (2005).
8. Matthess Y. Conditional inhibition of cancer cell proliferation by tetracycline-responsive, H1 promoter-driven silencing of PLK1. Oncogene. 24:2973–2980, (2005).
9. Thompson J.D. Improved accumulation and activity of ribozymes expressed from a

tRNA-based RNA polymerase III promoter. Nucleic Acids Res. 23: 2259–2268, (1995).

10. Ilves H. Retroviral vectors designed for targeted expression of RNA polymerase III-driven transcripts: a comparative study. Gene, 171: 203–208, (1996).
11. Boden D. Promoter choice affects the potency of HIV-1 specific RNA interference. Nucleic Acids Res., 31: 5033–5038, (2003).
12. Hassani Z. A hybrid CMV-H1 construct improves efficiency of PEI-delivered shRNA in the mouse brain. Nucleic Acids Res., 35: e65, (2007).
13. Nie L., Das T.M., Wang Y., Su Q., Zhao Y., and Feng Y. Regulation of U6 Promoter Activity by Transcriptional Interference in Viral Vector-Based RNAi. Genomics Proteomics Bioinformatics., 8(3): 170–179, (2010).
14. Su J., Zhu Z., Xiong F., and Wang Y. Hybrid Cytomegalovirus-U6 Promoter-based Plasmid Vectors Improve Efficiency of RNA Interference in Zebrafish. Marine Biotechnology, 10: 511–517, (2008).
15. Katrekar D., Chen G., Meluzzi D., Ganesh A., Worlikar A., Shih Yu-Ru., Varghese S., and Mali P. In vivo RNA editing of point mutations via RNA guided adenosine deaminases. Nature Methods, 16(3): 239–242, (2019)
16. Koyama M., Ishida T., Horiike K., Nozaki M., & Shimada M. Urate Oxidase Activity and Copper Content in the Liver of Macular Mutant Mouse, a Model Animal for Human Congenital Copper Deficiency, Menkes' Kinky Hair Disease. J. Nutr. Sci. Vitaminol, 37: 601-609, (1991)
17. Katsura T., Kawasaki H., Yamano T., and Himada M. Copper contents and pathological changes in various organs of macular mouse. Congenital Anom., 28: 85-92, (1988)

List of publications

1. **Sonali Bhakta,** Md Thoufic Anam Azad and Toshifumi Tsukahara Genetic code restoration by artificial RNA editing of Ochre stop codon with ADAR1 deaminase. Protein Engineering, Design and Selection, 31(12), 471–478 (2018).

2. **Sonali Bhakta** and Toshifumi Tsukahara. Artificial RNA editing with ADAR for genetic code restoration. Current gene therapy, 20(1), 44-54 (2020).

3. **Sonali Bhakta**, Matomo Sakari and Toshifumi Tsukahara. RNA editing of BFP, a point mutant of GFP, using artificial APOBEC 1 deaminase to restore the genetic code (Submitted-Scientific reports).

4. Md Thoufic Anam Azad, **Sonali Bhakta** and Toshifumi Tsukahara. Site-directed RNA editing by adenosine deaminase acting on RNA (ADAR1) for correction of the genetic code in gene therapy. Gene Therapy, 24, 779–786 (2017).

List of conferences attended (Domestic and International)

1. Poster: **Sonali Bhakta,** Md Thoufic Anam Azad and Toshifumi Tsukahara. Genetic code restoration by ADAR1 in Ochre stop codon. 19th general meeting of RNA society of Japan (2017).

2. Poster: Md Thoufic Anam Azad, **Sonali Bhakta** and Toshifumi Tsukahara. Comparative activity of Adenosine Deaminase acting on RNA (ADARs) isoforms for correction of Genetic code in gene therapy. 19th general meeting of RNA society of Japan (2017).

3. Oral: Toshifumi Tsukahara, Md Thoufic Anam Azad and **Sonali Bhakta.** Site-Directed RNA Editing by an Artificial Enzyme System. The 1st JAIST World Conference (2018).

4. Poster: **Sonali Bhakta,** Md Thoufic Anam Azad and Toshifumi Tsukahara. Artificial RNA editing in Ochre (TAA) stop codon for restoration of genetic code by using ADAR1 deaminase. The 1st JAIST World Conference (2018).

5. Poster: **Sonali Bhakta,** Md Thoufic Anam Azad and Toshifumi Tsukahara. Restoration of genetic code by artificial deaminase ADAR1 in Ochre (TAA) stop codon. JAIST Japan India Symposium, 2018.

6. Poster: **Sonali Bhakta** and Toshifumi Tsukahara. RNA editing in BFP (derived from point mutated GFP) by using artificial APOBEC1 deaminase for genetic code restoration. JAIST Japan India Symposium, 2019.

7. Oral: **Sonali Bhakta** and Toshifumi Tsukahara. Artificial RNA editing in (BFP) Blue Fluorescence Protein (derivative of GFP) by using APOBEC 1 deaminase for restoration of genetic code. BIT's 9th Annual congress of Molecular and Cell Biology, Singapore (2019).

8. Poster: **Sonali Bhakta** and Toshifumi Tsukahara. Genetic code restoration by using APOBEC 1 artificial cytidine deaminase. 42nd Annual meeting of Molecular Biology Society of Japan, Fukuoka (2019).

Awards

1. Monbukagakusho Scholarship (MEXT), 2017-2020 for Doctoral study.
2. JAIST foundation research grant, 2019 for attending conference in Singapore.
3. Best poster Award, 2018 in JAIST World Conference (JWC), JAIST, Japan.

Acknowledgement

This research entitled "**Artificial deaminase system for restoration of genetic code**" was performed at Tsukahara Laboratory, Bioscience and Biotechnology area, Graduate School of Advanced Science and Technology, Japan Advanced Institute of Science and Technology (JAIST), in the time period of October 2017 to September 2020.

I would like to express my sincere gratitude to ***Professor Dr. Toshifumi Tsukahara*** sensei of JAIST for his continuous guidance, suggestions, constructive criticism and kind support throughout my doctoral course. His motivating and cooperative nature helped me to work freely and broaden the research scope in various fields. The accomplishments during my Ph. D study would not be possible without his cooperation, encouragement and valuable suggestions.

I am also grateful to ***Associate Professor Dr. Takumi Yamaguchi*** sensei for his valuable guidance as my second research supervisor. Discussions and suggestions from him helped me a lot throughout the pathway of completing my thesis.

I would like to convey my grateful acknowledgement to my minor research supervisor ***Professor Dr. Kenzo Fujimoto*** sensei for giving me an opportunity to carry out the minor research or subtheme experiments in his laboratory and his valuable research guidance without which this study wouldn't be completed.

I am also very much grateful to ***Dr. Shigetaka Nakamura*** sensei, Assistant Professor, Area of Bioscience and Biotechnology, School of Advanced Science and Technology, for guiding me and constantly supervising my work. It has been a lifetime experience for me to learn the new techniques under his guidance. The author is highly pleased to express her deepest sense of appreciation, whole hearted gratitude to Dr. Shigetaka Nakamura sensei, for his cooperation and help during the research work.

I also express my indebted gratitude to ***Dr. Matomo Sakari*** Sensei, for helping and motivating me always through his valuable discussions on the topics regarding experimental methods and the data analysis.

I would like to give my sincere thanks to Saifullah, Jiarui Li, Ruchika Mishra, John Munene, Tetsuto Tohama, Wenhao Jin, Mitsuki Furuya, Junling Mo, Heeraman, Tomoko Onishi, Zhaoyuan Dai, Chisato Okhudaira, Tanaka san, Saitake san for their constant support and motivation during my whole PhD life in JAIST.

I am also very much thankful to Dr. Md. Thoufic Anam Azad and Dr. Umme Qulsum for their cordial guidance towards my PhD research, valuable suggestions and also cooperation. Those suggestions have really helped me to design my experimental plan and also to execute those.

I sincerely thank the Bangladeshi community in JAIST for encouraging me all the time and letting me to enjoy my life in JAIST. Apart from that I specially thank Dr. Aniruddha Nag for his constant support and being always there in my need.

Finally, I have no words to express my sincere gratitude and respect to my beloved parents Swapan Kumar Bhakta and Taposhi Bhakta, for flourishing love, mental support, cooperation and motivation all the time. I believe in the strength of my parent's blessings and without that this research wouldn't be possible.

Also, in my family, I am thankful to my elder brother Er. Sumon Kumar Bhakta, Sister-in-law Dr. Dipika Pramanik Bhakta and niece Shindhuza Rajashree Bhakta Priyonkori for all the support, blessings and love to me. I am thankful to my uncle Late Tapon Kumar Bhakta and Aunt Sheuli Bhakta, also to my younger brothers Shapath Kumar Bhakta, Aritra Bhakta and Borno Bhakta for their tremendous mental support and encouragement.

September, 2020 **Sonali Bhakta**
Japan Advanced Institute of Science and Technology