Chemically Modified guideRNAs are Resilient and Allow for Precise and Efficient Targeted A-to-I and C-to-U Editing

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Ngadhnjim Latifi
aus Vushtrri, Kosovo

Tübingen
2024

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen

Tag der mündlichen Prüfung	02.05.2024
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatter/-in:	Prof. Dr. Thorsten Stafforst
2. Berichterstatter/-in:	Prof. Dr. Ralf-Peter Jansen

Danksagung

Mein tiefer Dank gilt Prof. Dr. Thorsten Stafforst für die Möglichkeit ein so interessantes und sehr facettenreiches Thema bearbeiten zu können. Ich bedanke mich ferner für die Betreuung mit zahlreichen Ideen und Ratschlägen und dass er mir stets genügend Freiheiten und die Möglichkeit gab mich an der Chemie zu versuchen.
Ich danke auch Prof. Dr. Ralf-Peter Jansen recht herzlich für die Tätigkeit als Berichterstatter für diese Arbeit und die äußerst hilfreichen Vorschläge.

Dëshiroj të falënderoj nga zemra prindërit e mi, Vehbi dhe Hamijet Latifi. Përkushtimi dhe vullneti juaj janë frymëzim, dashuria dhe mbështetja juaj themeli që kanë bërë të mundur këtë vepër, të cilën jua dedikoj me gjithë zemër.
Për më tepër, dua të falënderoj motrat e mia Valbona Scharfenberg dhe Fjolla Latifi. Si dhe vëlllaun tim Vullnet Latifi. Ju vazhduat dhe hapët rrugën që unë thjesht duhej të ndiqja. Ju ishit shembuj dhe mentorë dhe për këtë ju falënderoj nga zemra.
Gjithashtu do të doja të falënderoja përzemërsisht partneren time Corina Popp dhe mikun tim më të mirë Marco Liewald, të cilët më mbështetën pa u lodhur në çdo rast, nuk u larguan kurrë nga unë dhe më shoqëruan në këtë rrugë me durim të pafund.

Des Weiteren möchte ich mich bei allen Mitgliedern der AG Stafforst für den tollen, anregenden und sehr lehrreichen Austausch während dieser Zeit bedanken. Ein besonderer Dank geht an die Mitglieder von „Los Cuatro Chemigos“ Anna Imrich, Oliver Heß und Alfred Hanswillemenke für eure Mühe und Geduld mit einem Chemieneuling. Ein weiterer besonderer Dank geht an Carolin Fruhner, Karthika Devi Kiran Kumar, Tobias Merkle und Philipp Reautschnig für eure Unterstützung und die zahlreichen Diskussionen zu interessanten Konzepten. Ferner bedanke ich mich ganz herzlich bei Yvonne Füll und Daniel Hofacker für die vielen tollen Gespräche und Ratschläge bei einem Kaffee oder einer schier unlösbaren Programmieraufgabe.

Nicht zuletzt möchte ich mich ganz herzlich bei Aline Mack und Clemens Lochmann bedanken, die mit Eifer, Fleiß und Zuverlässigkeit einen wertvollen Beitrag zu dieser Arbeit geleistet haben.

Table of Contents

Figures

Abbreviations

2’F	2’ fluoro
2’OMe	2’O methylation
A	context: adenine/adenosine, alanine
aa	amino acid
ACF/A1CF	APOBEC1 complementation factor
ActB	β-actin
ADAD	adenosine deaminase acting on DNA
ADAR	adenosine deaminase acting on RNA
ADAT	adenosine deaminase acting on tRNA
AID	Activation induced deaminase
AMPA	α-amino-3-hydroxy-5-methylisoxazol-4-propionic acid
APOBEC	Apolipoprotein B mRNA editing complex
APOE	apolipoprotein E
ASO	antisense oligo nucleotide
C	context: cytosine/cytidine, carbon, cysteine
Cas	CRISPR associated
cDNA	complementary/coding DNA
CIRTS	CRISPR-Cas-Inspired RNA Targeting System
CRISPR	clustered regularly interspaced short palindromic repeats
CTNNB1	β-catenin
CURE	C>U RNA Editing
D	aspartic acid/aspartate
DD	deaminase domain
DIC	N,N’-diisopropylcarbodiimide
DIPEA	N,N-diisopropylethylamine
DMSO	dimethyl sulfoxide
DNA	desoxyribonucleic acid
DRADA	dsRNA adenine deaminase
ds	double-stranded
dsRAD	dsRNA adenine deaminase
dsRED1/2	dsRNA specific editase1/2
E	glutamic acid

EDCI	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide
F	phenyl alanine
G	context: guanine/guanosine, glycine
GAPDH	glyceraldehyde 3-phosphate dehydrogenase
GluR	glutamate receptor
gRNA	guideRNA
GUSB	β-glucuronidase
H	context: hypoxanthine, histidine
hAGT	human O^6-alkylguanine-DNA alkytransferase
HEK	human embryonic kidney
HIV	human immunodeficiency virus
I	context: inosine, isoleucine
IgM,-A,-E,-G,	immunoglobulin M, A, E, G
KO	knock-out
LEAPER	leveraging endogenous ADAR for programmable editing of RNA
LNA	locked nucleic acid
M	methionine
mApo1S	murine APOBEC1-SNAP
mAPOBEC1	murine APOBEC1
mRNA	messenger RNA
N	context: nitrogen, asparagine, any nucleotide (A,T,U,G,C)
NHS	N-hydroxysuccinimide
O	oxygen
ORF	open reading frame
P	proline
pDNA	plasmid DNA
PO/PS	phosphate/phosphorothioate
PPIB	peptidyl-prolyl cis-trans isomerase B
PRPSAP2	phosphoribosyl pyrophosphate synthetase associated protein 2
Q	glutamine
R	arginine
RBD	RNA-binding domain
REPAIR	RNA Editing for Programmable A to I Replacement

RESCUE	RNA Editing for Specific C to U Exchange
RESTORE	recruiting endogenous ADAR to specific transcripts for oligonucleotide-mediated RNA editing
REWIRE	RNA-editing with individual RNA-binding enzyme
RNA	ribonucleic acid
RNAi	RNA interference
RNase	ribonuclease
S	context: serine, Svedberg
SDHB	succinate dehydrogenase B
siRNA	small interfering RNA
ss	single-stranded
STAT1/STAT3	signal transducer and activator of transcription 1/3
SV40	Simian virus 40
T	context: thymine/thymidine, threonine
TCF/LEF	T cell factor/lymphoid enhancer factor
TMEM109	transmembrane protein 109
U	uracil/uridine
UTR	untranslated region
V	valine
W	tryptophane
Y	tyrosine

Abstract

Deamination of adenosines and cytidines is mediated by the ADAR and AID/APOBEC families. The former family deaminates adenines in double-stranded RNA molecules leaving behind inosine, which acts as guanosine and is as such interpreted by ribosomes. The AID/APOBEC family deaminates cytidines primarily in DNA molecules but three of them were described to edit RNA. The most prominent of those three is APOBEC-1, which by deamination of cytidine creates a uridine leading to a Stop codon formation and is therefore responsible for the intestinal switch from ApoB-100 to ApoB-48 and thus substantially impacts lipid metabolism. Given the significance and potential of those processes, extensive efforts have been made to control and employ them. For A-to-I editing, multiple applications have been developed that either rely on endogenous levels of ADARs or on an over-expressed fusion of their deaminase domains and a protein promoting translocation to the target site. For both classes of applications, almost all tools rely on a guideRNA for flagging the target site by creating a double-strand with the target adenosine in a mismatch with cytidine and bearing some kind of structure for recruitment of the editing enzyme. Some of those tools rely on synthetic guideRNAs bearing common chemical modifications for stability (e.g. SNAP-ADAR) and some rely on plasmid-encoded guideRNAs. For targeted C-to-U editing, all reported tools rely on over-expressed fusion proteins with either an endogenous cytidine deaminase (e.g. APOBEC-1/-3A) or mutated ADAR2 deaminase domain. While for targeted A-to-I editing tools vary in efficacy but broadly show high-level editing across a multitude of targets, tools for targeted C-to-U editing suffer from either low editing levels and/or limited codon scope. In this work we further advance both branches of the field. By expanding chemical modification patterns of guideRNAs for SNAP-ADAR-mediated A-to-I editing we created a design that does not just retain almost full stability in extremely hostile environment up to a week of incubation but also allows for naked and carrier-free delivery to cells for editing of an endogenous transcript. In addition, we successfully augmented the highly potent SNAP-ADAR platform by HALO-tag. In fact, with HALO-tag we did not just provide an alternative to SNAP-tag, but we successfully recruited different effectors to different target sites simultaneously. This way, we ultimately employed two enzymes for concurrent A-to-I and C-to-U editing, the latter of which we achieved by fusion of SNAP-tag with the endogenous cytidine deaminase APOBEC-1 named mApo1S. However, erratic and low activity of mApo1S prompted us to explore other possibilities and we thus considered

RESCUE-S, a fusion of a mutated ADAR2 deaminase domain, capable of cytidine deamination, with a Cas13 protein. By exchanging Cas13 with the SNAP-tag we created SNAP-CDAR-S, a versatile tool with maximum programmability, high activity, and accuracy, which provides a valuable alternative to other current tools. In conclusion, the work presented here paves the way for a future multifaceted platform for eliciting various potentially multiplexed effects on target RNAs, which due to compound stability could even be applied to specific tissues or organs in complex living organisms.

Zusammenfassung

RNA-Editierung beschreibt einen natürlichen zellulären Vorgang, bei dem Adenosine oder Cytidine innerhalb eines RNA-Moleküls desaminiert werden. Dabei entsteht bei ersterem, Inosin, das sich wie Guanosin verhält und als solches auch von zellulären Proteinen, wie der Translationsmaschinerie erkannt wird. Dabei können unter anderem innerhalb einer kodierenden Sequenz Punktmutation von Adenosin nach Guanosin entstehen. Dieser Vorgang wird in der Zelle von Proteinen der ADAR-Familie ausgeführt. Die Desaminierung von Cytidin nach Uridin hat in ApoB ihr bekanntestes Beispiel. Hierbei führt die von APOBEC-1 ausgeführte Desaminierung dazu, dass das ApoB-100 Protein im Darm als verkleinertes ApoB-48 Protein vorzufinden ist. Neben APOBEC-1 wurden lediglich APOBEC-3A und -3G als Cytidin-desaminierende Proteine von RNA beschrieben. Die anderen Familienmitglieder der AID/APOBEC-Familie desaminieren für gewöhnlich Cytidine innerhalb von DNA-Molekülen. Das Potential, das diese Prozesse bergen führte zur Entwicklung zahlreicher Ansätze für eine zielgerichtete Anwendung. Für die A-nach-I-Editierung entstanden Ansätze, die entweder endogen exprimierte ADAR zielgerichtet rekrutieren oder aber von Fusionsproteinen Gebrauch machen, die aus der Desaminasedomäne der ADARs besteht und einem Proteinanteil, der die Translokation zur Zielsequenz vermittelt. Im Gegensatz dazu, wurden für die C-nach-U-Editierung ausschließlich Fusionsprotein entwickelt, die auf der katalytischen Aktivität von entweder endogenen Desaminasen (APOBEC-1, -3A, -3G) oder aber mutierten ADAR-Desaminasedomänen basieren. Während die Plattformen für A-nach-I Editierung weitestgehend effektiv sind, leiden die entwickelten Cytidindesaminasen entweder unter geringer Effektivität und/oder stark eingeschränktem Spektrum and zu editierenden Codons. Daher liegt das Bestreben dieser Arbeit darin das Feld der zielgerichteten RNA-Editierung anhand der nötigen Stellschrauben zu avancieren. Auf der einen Seite können wir hier zeigen, dass das SNAP-ADAR-System gängige chemische Modifikationen der guideRNA, die bereits in großem Umfang in anderen Oligonukleotideanwendungen Verwendung finden, nicht nur toleriert, sondern, dass diese sogar die Editierung verbessern können. Auf der anderen Seite können wir zeigen, dass diese chemischen Modifikationen auch zu einer stark erhöhten Stabilität der guideRNAs in extrem unwirtlichen Medien führen. Darüber hinaus ist es uns hier gelungen ein Modifikationsmuster zu entwickeln, das uns erlaubt in Zellkultur auf endogenen Transkripten zielgerichtet A nach I zu editieren, ohne auf Reagenzien angewiesen zu sein, um die guideRNAs in die Zelle zu schleusen.

Ferner ist es uns gelungen die SNAP-ADAR-Plattform um ein weiteres Enzym zur Interaktion mit der guideRNA zu erweitern. Somit haben wir mit diesem HALO-tag nicht nur eine Alternative zum SNAP-tag, sondern können auch unterschiedliche Effektoren zur Manipulation von RNA unabhängig voneinander und, aufgrund der hohen Substratspezifität, sogar gleichzeitig nutzen. Folglich ist es uns gelungen gleichzeitig A nach I und C nach U zu editieren. Letzteres gelang durch das Fusionieren von SNAP-tag mit der endogenen Cytidindesaminase APOBEC-1, welche jedoch aufgrund von mangelnder Kontrollierbarkeit eine suboptimale Plattform darstellte. Daraufhin bemühten wir SNAP-tag um eine bessere Plattform für zielgerichtete C nach U Editierung zu erschaffen, indem wir es mit einer mutierten ADAR2 Desaminasedomäne fusionierten, welche Cytidine als Substrat für Desaminierungen akzeptiert. Innerhalb des Cas13-basiertem RESCUE-S-System wurde es als kontrollierbar und mit einem breiten Spektrum an editierbaren Codons beschrieben. Indem wir Cas13 mit dem SNAP-tag austauschten, erschufen wir SNAP-CDAR-S, ein Enzym mit maximaler Programmierbarkeit, hoher Effektivität bei ausgezeichneter Genauigkeit und sehr breitem Wirkspektrum. Ähnlich wie bei SNAP-ADAR, ermöglichte das Zusammenspiel mit chemisch modifizierten guideRNAs eine hohe Editierungsausbeute von schwer zu editierenden Codons ohne an Genauigkeit einzubüßen. Zusammengenommen können die hier vorgelegten Ergebnisse den Grundstein einer vielseitigen Plattform legen, die es erlauben könnte, verschiedene Arten von RNA-Modifikationen zeitgleich, effizient und aufgrund von hoher Stabilität potenziell sogar zielgerichtet auf Geweben oder Organen innerhalb komplexer Lebewesen anzuwenden.

Publications and other collaborative work

Publication 1 (published)

Precise and efficient C-to-U RNA base editing with SNAP-CDAR-S

Ngadhnjim Latifi, Aline Maria Mack, Irem Tellioglu, Salvatore Di Giorgio and Thorsten Stafforst, *Nucleic Acids Research*, 2023, *51.15,* e84

Personal contribution

Chemical synthesis of linker bearing the substrate for SNAP-tag. Design of all NH-guideRNAs together with Thorsten Stafforst and subsequent conjugation with snap-linker. Design, plan, and analysis of all editing experiments together with Thorsten Stafforst (transcriptome sequencing experiment not included). Performance of all editing experiments except for the ones on eGFP transcript with the mApo1S constructs bearing different localization tags. Co-supervision of Aline Maria Mack, who performed editing experiments on eGFP transcript with mApo1S bearing localization tags. Performance of the Western Blot and β-catenin-Luciferase Assays. Performance of transfection and sample preparation for transcriptome sequencing experiment. Analysis and visualization of transcriptome sequencing data after QC, adapter trimming, mapping to genome, and editing event call (performed by Irem Tellioglu and Salvatore Di Giorgio) together with Thorsten Stafforst. Creation of schematics, tables and visualization of all data. Preparation of all supporting data and extended Methods & Materials. Preparation of Methods & Materials of main manuscript and contribution to description and discussion of results. Contributions to implementation of Reviewer's requests.

Publication 2 (published)

Harnessing self-labeling enzymes for selective and concurrent A-to-I and C-ot-U RNA base editing

Anna S. Stroppel, **Ngadhnjim Latifi**, Alfred Hanswillemenke, Rafail Nikolaos Tasakis, F. Nina Papavasiliou and Thorsten Stafforst, *Nucleic Acids Research*, 2021, *49.16,* e95

Personal contribution

Design, plan, and analysis of editing experiments for benchmark between mApo1S and RESCUE together with Thorsten Stafforst and Anna S. Imrich (née Stroppel). Execution

of editing experiments for benchmark between mApo1S and RESCUE. Execution of editing experiment and preparation of samples for transcriptome sequencing concurrent A-to-I and C-to-U editing. Analysis of transcriptome sequencing data after QC, adapter trimming, mapping to genome, and editing event call (performed by Rafail Nikolaos Tasakis) together with Thorsten Stafforst and Anna S. Imrich (née Stroppel).

Publication 3 (published)

CLUSTER guideRNAs enable precise and efficient RNA editing with endogenous ADAR enzymes in vivo

Philipp Reautschnig, Nicolai Wahn, Jacqueline Wettengel, Annika E. Schulz, **Ngadhnjim Latifi**, Paul Vogel, Tae-Won Kang, Laura S. Pfeiffer, Christine Zarges, Ulrike Naumann, Lars Zender and Thorsten Stafforst, *Nature Biotechnology*, 2022, *40.5*, 759-768

Personal contribution

Performing of animal experiments together with Philipp Reautschnig, Jacqueline Wettengel, and Tae-Won Kang. This entailed, preparation of mice for injection (injection performed by Tae-Won Kang), euthanasia of mice at endpoint, dissection, and liver extraction. Subsequent total RNA isolation and sample preparation for Sanger sequencing. Western Blot for detection of ADAR1 protein from cell stably expressing ADAR1 under doxycycline-induction.

Collaborative work

Targeted A-to-I editing with chemically modified guideRNAs (section 2.2)

Personal contribution

Design of all NH-guideRNAs together with Thorsten Stafforst except for guideRNAs #1-8. Design, plan, and analysis of all editing and stability experiments together with Thorsten Stafforst. Conjugation of snap-substrate to all guideRNAs in section 2.2.1. Execution of all editing and stability experiments in section 2.2.1. Creation of all cell lines stably and constitutively expressing mMeCP2-eGFP variants, respectively. Co-supervision of Clemens Lochmann who performed all editing and stability experiments, as well as snap-substrate conjugation to guideRNAs in section 2.2.2 and 2.2.3 for his Bachelor's thesis. All stability experiments and conjugation of snap-substrate to guideRNAs in section 2.2.4. Co-

supervision of Clemens Lochmann who performed all editing experiments in sections 2.2.4 and 2.2.5 during his laboratory rotation. Creation of tables and visualization of all data as well as preparation of Methods and Materials section.

1. Introduction

1.1. RNA base editing

1.1.1. Discovery

Lipid metabolism is a crucial part of energy generation, retaining cell membrane integrity, and many other functions.[1] Therefore, transport of ingested lipids in an aqueous milieu such as blood is pivotal. Consequently, this is achieved by lipoproteins acting as vehicles for transport through the venae system. An integral part of these vehicles is apolipoprotein B (ApoB).[2, 3] Two different forms of ApoB can be found in mammals, the full-length protein (ApoB-100) and one of approximately 48 % of its size (ApoB-48). ApoB-100 constitutes an important part of very low density lipoproteins and later the major part of the low density lipoproteins (LDL) and is expressed in multiple tissues with levels varying upon developmental stages of fetal tissue.[4, 5] However, in human liver, ApoB-100 constitutes the only expressed ApoB throughout all stages.[4-6] In fetal intestinal tissue, ApoB-100 is also predominantly expressed, which however gradually changes to ApoB-48 in adult small intestine. Here, it is an integral part of chylomicrons.[6] The existence of these two protein were subject of extensive research.[2, 6-8] Eventually, sequence analysis of the coding desoxyribonucleic acids (cDNA) of ApoB of the small intestine revealed that the sequence coding for the amino (N)-terminus of those two protein species were identical.[2, 7-13] However, at codon 2153 coding for glutamine (Q) within the ApoB-100 transcript, the cytidine (C) was replaced by a uridine (U) within the ApoB-48 transcript.[2, 7-13] This point mutation that was exclusively found in the messenger ribonucleic acid (mRNA) causes an in-frame stop codon formation resulting in a truncated protein.[2, 7-13]

More or less simultaneously, the first case of adenosine (A) to inosine (I) editing within an mRNA molecule was reported when investigating why gene knock-down in *Xenopus laevis* oocytes by hybridization of target mRNA with a complementary oligonucleotide could not be transferred to embryonic cells.[14, 15] Examination unveiled that the oligo-mRNA duplex was partially unwound in a time dependent manner, a phenomenon also found in a variety of mammalian cells.[14-17] Further investigation revealed that this irreversible separation of the two strands was due to deamination of adenosines resulting in inosine. Since inosine behaves like guanosine (G) and thus Whatson-Crick base pairs with cytidine rather than uridine, the duplex of the mRNA was weakened, leading to the unwinding phenomenon. However, besides base pairing, inosine is also interpreted a guanosine by the translation

machinery and thus ultimately leads to A-to-G point mutation within a coding sequence and is thus accompanied by the corresponding amino acid change.[18-21]

1.1.2. Adenosine deamination on RNA

Adenosine deamination of RNA is catalyzed by more than one enzyme, which due to their homology are collectively called adenosine deaminases that act on RNA (ADAR).[22] Apart from some unique domains, all ADARs share similar traits with N-terminal double-stranded RNA binding domains (dsRBDs) followed by a carboxy (C)-terminal deaminase domain. Intriguingly, only ADAR1 and ADAR2 are responsible for reported A-to-I RNA editing, ADAR3 does not possess any so far reported catalytic activity.[23-25]

ADAR1 initially termed dsRNA adenine deaminase (DRADA[26]/dsRAD[27]) was first isolated from organ extracts of different species. It contains three dsRBMs in the N-terminus with the third one bearing a nuclear localization signal (NLS). Expression is ubiquitous and triggered from different promoters leading to two ADAR1 species of different sizes. ADAR1 p110, is constitutively expressed and mainly present in the nucleus but can shuttle to the cytoplasm. This is mediated by proteins of the exportin family and regulated by binding of dsRNA to the dsRBDs. Return to the nucleus is mediated by interaction of transportin 1 with the NLS at the 3rd dsRBD, which in turn is inhibited by binding of dsRNA. ADAR1 p150 expression is induced by interferon, forming an ADAR1 isotype with an elongated N-terminus containing a Zα domain with a nuclear export signal (NES).[24-32]

ADAR2 (ADARB1) was first found in rat brain tissue as an enzyme with only two dsRBDs named dsRNA specific editase1 (dsRED1[33]). The dsRBDs show homology to ADAR1's dsRBDs one and three, respectively. Consequently, an NLS can also be found on the N-terminus locating ADAR2 in the nucleus. Binding of karyopherin subunit α 1 & 3 to the arginine rich NLS causes the predominant nucleolar localization of ADAR2. Moreover, protein fate can be regulated by post-translational modification. That is, ubiquitinylation causes cytoplasmic degradation of the protein whereas phosphorylation of threonine (T) 32 leads to increased nuclear localization. Even though tissue expression is also ubiquitous, it is strongly expressed in brain tissue.[24, 25, 33, 34]

ADAR3 has high sequence homology to ADAR2. In the N-terminus it contains an arginine-rich site that is believed to be used as a single stranded RNA (ssRNA) binding motif. Different from the other ADARs, this family member is only expressed in cells of the

central nervous system. As of now, no deaminase activity was reported for this family member.[23-25, 35]

1.1.3. Cytidine deamination on RNA

Unlike A-to-I editing, cytidine deamination within ApoB mRNA is performed by a multidomain complex called "editosome".[36-45] The catalytically active moiety within this complex constitutes apolipoprotein B mRNA editing enzyme, catalytic polypeptide 1 (APOBEC-1) and belongs to the activation induced deaminase (AID)/APOBEC family consisting of elven members of cytidine deaminases. Although heterologous in their role, tissue distribution and substrate preference, they all have at least one Zinc (Zn^{2+})-dependent deaminase domain that bears a similar residue formation that can be found in cytidine deaminases across all domains of life and even ADARs.[26, 30, 43, 46-52] While APOBEC-2 and -4 have no reported cytidine deamination activity, the other members can elicit editing within double stranded DNA (dsDNA), single stranded DNA (ssDNA), and RNA and are of central importance for innate and adaptive immunity, and other cellular functions.[51, 53, 54] In this work, we will focus on the RNA editing APOBECs, in particular APOBEC-1, but the other members will be briefly introduced. The reader is further referred to literature cited here[51, 53-56] for further information on DNA editing AID/APOBEC family members.

AID is a pivotal factor of immune reactions and deaminates cytidines in an ssDNA context but can also target cytidines in a supercoiled dsDNA context.[57, 58] Over the course of an infection, B-cells are involved in production of immunoglobulin M (IgM) antibodies. Once activated, those B-cells switch the classes of the antibodies from IgM to IgA, IgE, or IgG. The underlying process for this is somatic hypermutation and class switch recombination. AID is involved and in part responsible for both mechanisms.[51, 56] What is more, AID and APOBEC-2, that is homologues in other species, are considered the ancestors of all other APOBECs, which themselves are proposed to result from gene duplication and diversification events over millions of years of evolution.[51, 52]

The largest family constitutes APOBEC3 that consists of seven members named APOBEC-3A, B, C, DE, F, G, and H. While APOBEC-3A, -3C, and -3H have one deaminase domain, the other members have two, of which however, only one seems to be the catalytically active one. All members are expressed in immune cells and constitute an antiviral defense. APOBEC-3DE, -3F, -3G, and -3H are reported to restrict human immunodeficiency virus 1 (HIV-1) infection in T-cells. HIV-1's viral infectivity factor (Vif) binds in turn to

domains within those APOBEC-3 species causing their ubiquitinylation and subsequent degradation. Moreover, with lower APOBEC-3A expression levels of macrophages as compared to monocytes, susceptibility of macrophages to HIV-1 infection increases. While all APOBEC-3 are specialized to edit cytidines within ssDNA molecules, they have the capacity to bind RNA.[51, 53-55] In fact, APOBEC-3A and -3G were even shown to be able to edit RNA. In monocytes RNA editing by APOBEC-3A is activated under hypoxic conditions or upon interferon induction. Interestingly, when over-expressed in human embryonic kidney (HEK) 293T cells, most of the target sites of APOBEC-3A can also be found in monocytes upon RNA editing activation. In addition, there is only a small degree of overlap between over-expressed APOBEC-3A and -3G hinting at individual and distinct substrate preference of both enzymes.[59-62]

APOBEC-1 is a 27 kDa protein with homologues in multiple species that perform ApoB mRNA editing.[63-66] In fact, APOBEC-1 is the obligatory catalytic component for ApoB mRNA editing, as its expression correlates with editing occurrence. Consequently, in an APOBEC-1-depeleted mouse model no ApoB mRNA editing was detected in the intestine or the liver but could be reconstituted by supplementation of APOBEC-1 as viral vector. However, APOBEC-1 depletion substantially impacted not only lipid metabolism but also brain homeostasis. Different from humans, mice lack APOBEC-3 and therefore rely on APOBEC-1 for editing in immune cells. In the brain, microglia, a type of macrophage, are responsible among other things for phagocytosis of pathogens, cell debris, and other proinflammatory effectors. Lack of APOBEC-1 severely reduces anti-inflammatory functions of microglia.[67-70]

Unlike the other family members, APOBEC-1 requires co-factors that assemble in an editosome to elicit ApoB mRNA editing.[36-45] The editosome is a protein complex with a sedimentation coefficient of 27 Svedberg (S) localized in the nucleus. However, proteins within that complex can also be found in a cytoplasmatic and editing inactive complex of 60S believed to serve as a reservoir for proteins involved in ApoB mRNA editing.[38-41]

The mode of action of the editosome has not been entirely elucidated, as contradictory results made a general assumption difficult. Initially, APOBEC-1 complementation factor (ACF/A1CF) was identified as a 64-65 kDa RNA binding protein that colocalized with APOBEC-1 in editing active 27S complex and was a sufficient supplement for ApoB mRNA editing *in-vitro*.[71-76] In a proposed model, ACF binds to the mooring sequence of the ApoB mRNA target site and interacts with two APOBEC-1 subunits to position the dimer for deamination. This was considered the minimal requirement for ApoB mRNA

editing.[71, 72] Other proteins in the complex such as KH-Type Splicing Regulatory Protein (KHSRP) or glycine-arginine-tyrosine-rich RNA-binding protein (GRY-RBP) and others were found to act as editing inhibitors by interaction with APOBEC-1 or ACF. [71, 77-83]

In accordance with evidence of ApoB mRNA editing being a nuclear event and editing activity correlating with nuclear localized 27S editosome, APOBEC-1 was shown to shuttle between nucleus and cytoplasm.[84-88] In fact, the N-terminus was found to be involved in nuclear localization but the exact residues responsible remained elusive. The same holds true for the C-terminus involved in the cytoplasmatic localization/cytoplasmic retention of APOBEC-1.[84-88] Discovery of ACF, which was also localized in the nucleus, then raised the question if ACF itself managed to drag APOBEC-1 to the nucleus, but contradicting results of different groups failed to completely clarify this issue.[88, 89] Nevertheless, subcellular localization of ACF greatly impacts editing and was found to be dictated by phosphorylation, which itself is influenced by dietary and hormonal factors.[90-92]

Besides localization, ApoB mRNA editing was shown to be regulated in further ways. Firstly, expression of APOBEC-1 is developmentally regulated.[93] Interestingly, in rats and mice, transcription is triggered by different promoters, one of which is responsible for expression in liver and other organs. In humans, APOBEC-1 gene lacks that promoter, explaining the absence of ApoB mRNA editing in human liver tissue.[94-96] Secondly, an alternative splice variant of APOBEC-1 was found that codes for a non-functional protein. In early stages of development this splice variant made up 90% of all APOBEC-1 transcript and was reduced to 50% in adult tissue.[97] This balance of non-functional and functional protein constitutes an important regulatory aspect, as promiscuous editing in both cell culture and animal model, leading to cancer development in the latter, was reported upon over-expression of APOBEC-1.[98-101] In fact, APOBEC-1 was indeed shown to edit DNA.[102, 103] Moreover, diet-dependent alternative splice variants of ACF that either bind to ApoB mRNA but fail to interact with APOBEC-1 and vice-versa were shown to regulate ApoB mRNA editing, as well.[104, 105]

Intriguingly, while APOBEC-1-depleted animals were still viable, ACF-knock-out (KO) proved detrimental in early embryonic stages, indicating implications going far beyond ApoB mRNA editing.[106] In 2017 a different group reported successful KO by delaying the onset of depletion to later stages of embryogenesis.[107] Strikingly, they did not just report no reduction but a slight increase in ApoB mRNA editing in absence of ACF suggesting that *in-vivo* ApoB mRNA editing does not rely on ACF despite its reported relevance for *in-vitro* complementation.[71-76, 107] An explanation for this had been found in 2014 by Fossat

and colleagues who reported RNA-Binding-Motif-47 (RBM47) as a 64 kDa protein to be of requisite importance for ApoB mRNA editing *in-vivo*.[108] In fact, for RBM47 as well, a depletion exhibited embryonic lethality for some mice. Viable animals showed hardly any ApoB-48 in intestinal tissue demonstrating its importance for ApoB mRNA editing. What is more, other discovered targets of APOBEC-1 also exhibited reduced editing.[109] While RBM47 was sufficient to supplement ApoB mRNA editing *in-vitro* in absence of ACF, it was also shown that it can interact with ACF, besides APOBEC-1. In contrast to the intestine, ApoB mRNA editing in hepatic tissue was not entirely impaired but only severely reduced upon RBM47-depletion. In fact, this tissue-specificity of editing supplementation by ACF and RBM47 was not limited to ApoB mRNA, as multiple reported APOBEC-1 editing sites exhibited altered editing depending on which complementation factor was missing.[110, 111]

1.1.4. Sequence requirements

Given the accuracy at which APOBEC-1 performs ApoB mRNA editing under physiological conditions, the target sequence was subject of extensive research and unveiled various causative elements for efficient editing.

The minimal sequence requirement varied upon experimental set up from 20 to 26 nt surrounding target C^{66666}, sequences that show perfect homology across multiple species capable of ApoB mRNA editing.[2, 13, 112-116] Computational analyses predicts a stem-loop structure formed by target C^{6666} flanking sequences while placing it at the 5' end of a loop (Figure 1A), which could also be confirmed by NMR analyses.[2, 13, 112, 116-118] The loop forming "spacer element" bearing target C^{6666} separates the 5' located "efficiency element" and the 3' located "mooring sequence" (5'-UGAUCAGUAUA).[38, 40, 117-120]

Mutational analysis of these elements revealed varying degrees of importance. On the one hand, mutations of the efficiency element or size variations of the spacer elements modulated editing but were largely tolerated. Interestingly, insertions of additional Cs within the spacer element resulted in their editing, while maintaining and sometimes even boosting on-target editing. In stark contrast to that, sequence integrity of the mooring sequence was an absolute requisite for editing. Mutations of all but one site (central guanine) either dramatically decreased or completely abolished editing.[117-119] In fact, the mooring sequence can support editing at an upstream C even when placed in a sequence

context outside of ApoB.[121] Moreover, using the mooring sequence as a probe an additional APOBEC-1-dependent C-to-U editing site was found in the transcript coding for

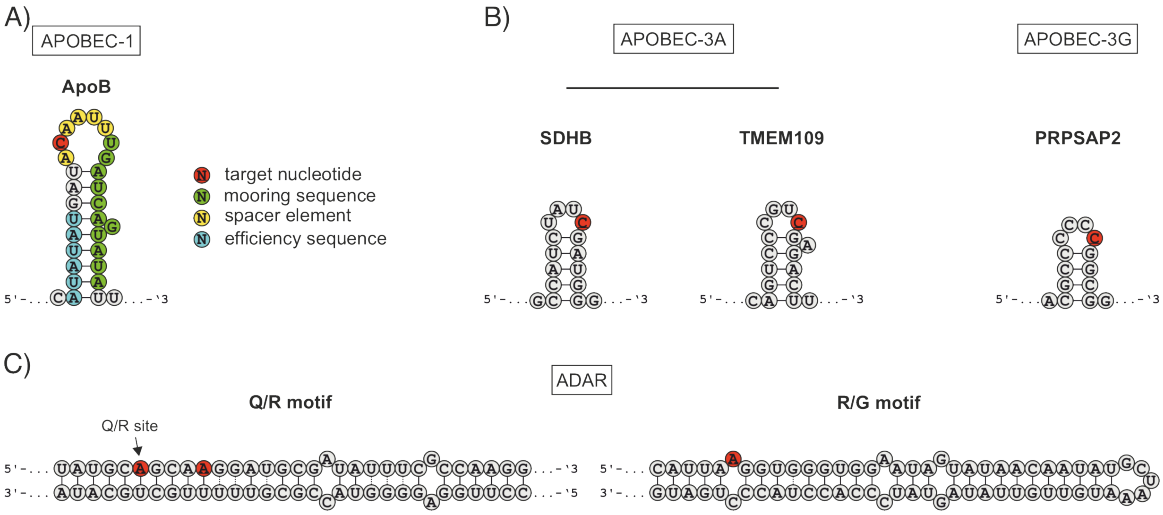


Figure 1: ***Secondary structures of endogenous targets of APOBECs and ADARs. A)*** *ApoB mRNA target predicted to form a secondary structure. Sequence elements have various levels of importance, the highest of which has the mooring sequence.* ***B)*** *Predicted secondary structures of APOBEC-3 targets succinate dehydrogenase B (SDHB), transmembrane protein 109 (TMEM109), and phosphoribosyl pyrophosphate synthetase associated protein 2 (PRPSAP2).* ***C)*** *ADAR-target sites are predicted to form secondary structures with intronic sequences. Besides, the declared target adenine, other sites can serve as ADAR substrates. In addition, mismatching the Q/R site's target adenine with a cytosine opens it up to editing by recombinant ADAR1, as it is the case for the R/G site. Secondary structures, and information adapted from [13, 21, 112, 116-118, 122-127].*

neurofibromatosis type I, which additionally showed sequence homology to ApoB mRNA's spacer and efficiency element.[128] Transcriptome-wide sequencing uncovered APOBEC-1-dependent mRNA editing in multiple transcripts, most of which were located in the transcripts' 3'untranslated regions (UTR) and all of which were located 2-5 nucleotides upstream of a variation of the ApoB mooring sequence.[109, 129] Nearest neighbor preference analysis revealed that both 5' and 3' APOBEC-1 prefers adenine or uracil with a slight inclination towards adenine at the 5' position. Accordingly, all found sequences were in an AU-rich context, similar to the ApoB target itself. There, 150 nt surrounding target C^{6666} contains 74% AU, while the rest of the entire transcript only contains 57%. In fact, placing ApoB minimal editing cassette in a GC-rich context decreased editing dramatically.[112-116, 120, 129, 130]

Besides AU-content and presence of a mooring sequence (or variation of it), secondary structure was also of relevance for increased editing outside of ApoB. In fact, it was shown that localization of the target cytidine inside the loop of a stem-loop structure and a mooring sequence within the stem influenced editing yield most positively.[130]

Accordingly, for APOBEC-3A and APOBEC-3G almost all targets were predicted to be inside the loop of a stem-loop structure formed by palindromic sequences flanking the target Cs (Figure 1B).[59-62] Indeed, mutations that impede stem-loop formation could severely reduce editing.[60] Moreover, APOBEC-3A prefers 5'-thymidine (T)CA/G contexts, while APOBEC-3G prefers a 5'CCN. Those nearest neighbor preferences were determined by residues within a loop of the deaminase domains (section 1.1.5).[59, 61, 131, 132]
A similar dependency on double-stranded structures was also observed for endogenous ADAR editing sites (Figure 1C). One of the first was a glutamine to arginine (R) change within the glutamate receptor (GluR) subunit of the α-amino-3-hydroxy-5-methylisoxazol-4-propionic acid (AMPA) receptor. Analysis of the mRNA showed that the editing site is within a double strand containing G:U wobble pairs that is formed by the exonic region and a downstream adjacent intronic region. Besides this Q/R site, the following glutamine (5'-CAA codon) was also edited, in addition to two observed editing hot spots further downstream.[21, 122-124, 126, 127] Other adenines present in the construct were not edited, contradicting the observation that ADAR edits adenines arbitrarily within a perfectly matched duplex.[14, 15, 18] However, this indicated that this duplex structure with wobbles, mismatches, and loops constitutes an element to steer ADARs to designated target sites.[21, 122-124, 126, 127] This was also reported for an editing site within exon 13 of GluR-B, -C, and -D inducing an arginine to glycine (G) mutation.[125] For this R/G site again an intronic sequence was discovered positioning the target A within a not perfectly matched double strand (Figure 1C). However, here the target A was in a mismatch with a C, which proofed to be a sticking point, as this R/G was moderately edited by both ADARs.[123-127] The Q/R site however was predominantly edited by recombinant ADAR2 and could only be edited by recombinant ADAR1 when the target A was put in a mismatch with a C.[33, 127] The requirement for an imperfect double-strand for editing was observed in even more targets.[133-135]
Similar to the APOBECs, ADARs also have preferences for nearest neighbor nucleotides. At the immediately 5' position both ADARs prefer U and A over C and G with the latter being the least favorable. Immediately 3' of the target A, both ADARs prefer G the most, while preference for the other nucleotides slightly diverge. ADAR1 prefers C and A equally and U the least and ADAR2 prefers C over U and A, which both are of equal preference.[136-138] Transcriptome-wide analysis of A-to-I editing events revealed that most of the sites were located in introns and 3'UTRs of genes, most of which belonged to the Alu elements, a

structured feature of primate mRNA transcripts. This is in accordance with the before found targets and further substantiates ADARs requirement of secondary structures for endogenous targeted editing.[139]

1.1.5. The Deaminase Domains

Upon cDNA isolation of APOBEC1 and the subsequent decoding of primary protein structure, functionally crucial residues were found in a formation that was known from cytidine deaminases of human, bacterial, and viral origin.[43, 46-49] In this motif, consisting of histidine (H), alanine (A)/valine (V), glutamate (E), separated by a number of amino acids (X_n) from proline (P) and two cysteines (C), looked as follows: $H(A/V)EX_{nn}PCX_2C$. Here, the histidine and the cysteines coordinate Zn^{2+}, which binds a water molecule. The glutamate then activates the water for a nucleophilic attack of the C4 of the target cytosine and shuttles a proton to the amino group forming an NH_3^+ leaving the reaction resulting in uracil.[140]

Most intriguingly, the same residues in a similar formation were found in both, the deaminase domains of the ADARs, and in the deaminase domains of their close relatives Adenosine deaminase actin on tRNA (ADAT).[24, 26, 30] Here again, deamination is carried out as a nucleophilic attack at adenine's C6 with the glutamate activating a water molecule bound to Zn^{2+} coordinated by histidine and two cysteines and shuttling a proton to the amino group. An inositol hexakisphosphate in vicinity of this deamination motif was shown to essential for function.[141]

Besides this glutamate, the target adenine was shown to interact with several residues of ADAR2 deaminase domain (Figure 2B) , among which was threonine 375 (T375).[142] More accurately, T375 interacts with the 2'-hydroxyl (2'OH) group of the ribose. However, most intriguingly, when a cytidine analog inhibitor was employed, structure overlay predicted a distortion as compared to when an adenine is deaminated leading to a clash of T375 with the ribose moiety of the inhibitor and thus giving a possible explanation as to why ADARs do not deaminate cytidines, despite a similar catalytic site.[141, 142]

For deamination, the target adenine is flipped out from the double strand, a mechanism that resembles the mode of action of DNA-methyltransferase.[143, 144] When flipped out, ADAR2 deaminase domain invades the double strand from the minor grove side with glutamate 488 (E488) flanked by glycines on either side (G487 & G489) and interacts with the orphan base left by the adenine flipping out and the hydroxyl group of the nucleotide immediately

5' of target A.[142] This interaction also explains ADAR's preference for pyrimidines opposite of the target adenine, as purines would cause spatial issues.[142]

What is more, this strand invasion motif elucidated the 5' nearest neighbor preference of ADARs. It had been shown that ADARs prefer uridine and adenine over cytidine and guanine 5' of the target adenine with guanine being detrimental for editing (Figure 2C).[136-138, 145] As strand invasion upon base flipping is carried out from the minor grove side, guanine's amino group, which points into the minor grove of the double strand clashes with the glycine flanking the glutamate and thus hampers deamination (Figure 2C, upper panel).

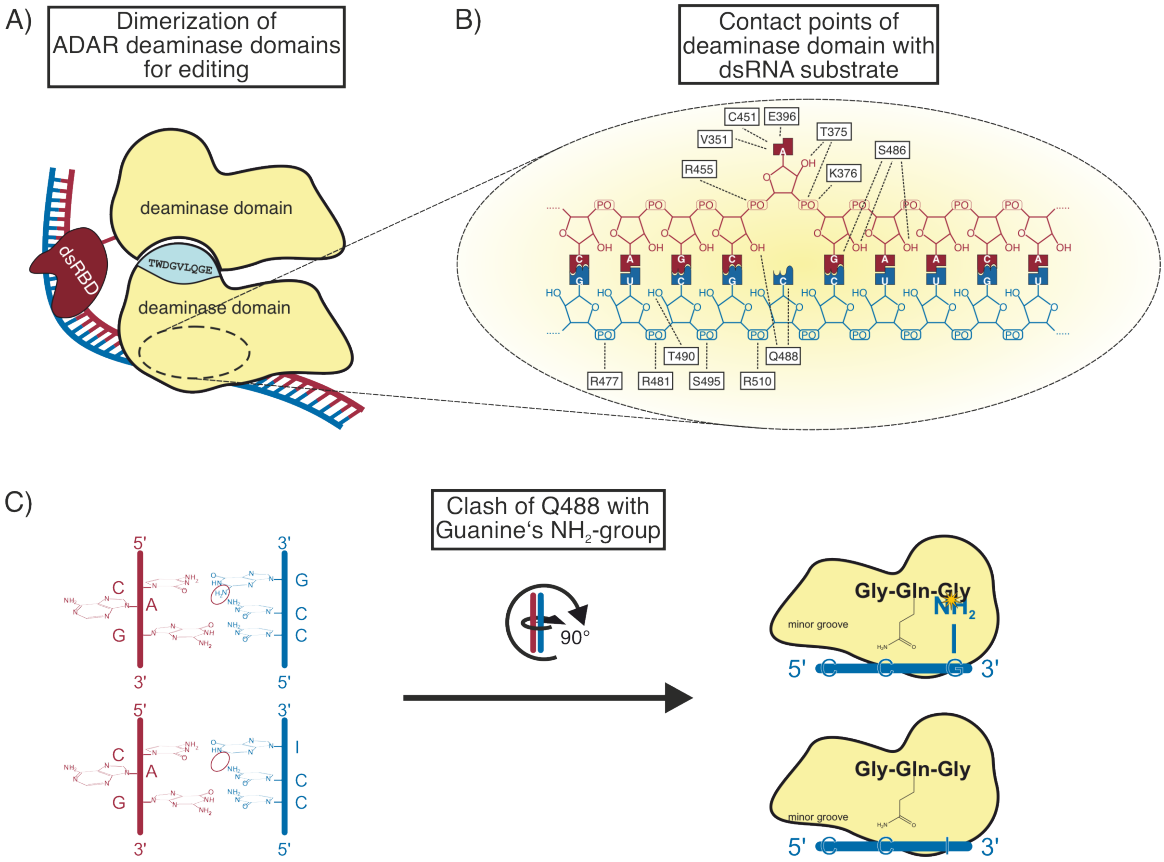


*Figure 2: **ADAR2 deaminase domain. A)** Editing is performed in as an asymmetric dimer. One ADAR2 deaminase domain binds to the target site for editing, while a second ADAR2 deaminase domain bind to an α-helix within the first deaminase domain. In addition, the dsRBD of the second ADAR2 binds to the dsRNA substrate. **B)** Contact points of the deaminase domain with the target site. The target A is flipped out of the double-strand into the catalytic pocket for deamination. The vacated space is taken up by invasion of either a glutamate (E) of a wildtype ADAR2 deaminase domain or a glutamine (Q) of a hyperactive mutant ADAR2 deaminase domain, which both interact with the orphan C and a ribose moiety of target A's neighboring nucleotide. **C) Left panel. Point of view: from the minor grove side.** Non-preference of a G 5' of the target A is determined by the strand invasion mechanism of the deaminase domain. Strand invasion by Q/E488 is performed from the minor grove side. Here, the NH2 of a G protrudes into the minor grove (left upper panel, red circle, protruding towards the point of view). Hypoxanthine (inosine's base) lacks that group (left lower panel, red circle). **Right panel. Point of view: from the backbone side of the dsRNA substrate.** The protrusion of guanine's NH2 group causes a clash with one of the glycines flanking Q/E488 and severely hampers deamination (right upper panel). Substituting guanine with hypoxanthine, which lacks that group does not lead to a clash (right lower panel). Information and schematic taken and adapted from [142, 146].*

This was confirmed by employment of inosine at that site, as hypoxanthine (inosine's base) lacks that amino group and thus does not cause the steric clash (Figure 2C, lower panel).[142] In contrast, when guanosine is located 3' of the target A, guanine's amino group interacts with serine 486 (S486) thus explaining this nearest neighbor preference.[142]

Besides flanking nucleotides, pH-dependency of E488 interaction with the orphan base proved to be a potent modulator of editing activity. Only under acidic conditions (i.e., pH lower than physiological 7.4) E488 side chain is protonated and is therefore capable of serving as a hydrogen bond donor to the orphan base for enhancement and stabilization of base flipping. In contrast, a glutamine on that position is already protonated at pH 7.4 and can readily serve as hydrogen bond donor explaining why the known E488Q (E1008Q in ADAR1) mutant is so much more active than its wildtype counterpart.[142, 145, 147, 148]

Apart from catalytic activity, the deaminase domain's involvement is more extensive. ADARs have been reported to dimerize for editing and crystal structure of a dimer of ADAR2 bearing only one dsRBD gave rise to a proposed model of an asymmetric dimerization (Figure 2A).[146, 149-151] While the deaminase domain of one ADAR binds to the target site for deamination, the dsRBD of another ADAR binds the RNA duplex. In addition, this ADAR's deaminase domain binds to an α-helix within the deaminase domain performing deamination. This α-helix is formed by residues 501-509 (1021-1029 in ADAR1) of which residues T501, tryptophane (W502), aspartate (D503), and G504 (T1021, W1022, D1023, G1024 in ADAR1) are conserved between ADARs of different species and are contacted by the deaminase domain with residues of its catalytic center. Indeed, mutations of this α-helix inhibited editing on multiple endogenous targets.[146]

Despite similar residue formation with respect to Zinc coordination, only a small part of structural features is conserved among ADARs and cytidine deaminases.[141] Consequently, the mechanism by which substrate is recognized and positioned by the APOBECs is remarkably different. As APOBECs do not have dsDNA/RNA binding domains, substrate binding is mediated by a groove lined with cationic residues for interaction with the backbone and aromatic residues for π-stacking with the bases leading to the catalytic center of the deaminase domains.[51, 152]

The AID/APOBEC family can be subdivided into members with one deaminase domain (AID, APOBEC-1, -2, -3A, -3C, -3H, and -4) and members with two (APOBEC-3B, -3DE, -3G, and -3F). For the latter, only the C-terminal deaminase domain exhibits editing activity, while the N-terminal one does not, despite containing necessary residues and

maintaining the super-secondary structure of active deaminase domains, consisting of alternating α-helices and β-sheets, which are folded in a way that the six α-helices surround the five β-sheets in the center.[51, 56]

The Zinc-coordinating residues lie at the N-termini of α-helices 2 and 3, respectively. Within this fold, intervening loops 1, 3, 5, and 7 make up a U-shaped groove, in which the substrate is bound and positioned for deamination. This U-shaped groove can fit the hairpin of the target sites (section 1.1.4) and resembles the structure of the tRNA adenine deaminase TadA.[51, 153]

Unfortunately, there is no reported crystal structure of APOBEC-1 bound to its substrate but given the conserved fold of the deaminase domains, crystal structures of APOBEC-3A and -3G bound to ssDNA substrate give valuable information that can be transferred to APOBEC-1. Besides Zinc-coordination, other residues within the groove have distinct features important for the reaction, substrate specificity and are in part conserved among cytidine deaminases. For instance, within the groove the target C is flipped into a small pocket and π-stacks with a tyrosine (Y) within APOBEC-3A (Y130) and APOBEC-3G (Y315) or a phenylalanine (F) within APOBEC-1 (F120).[56, 154, 155] Moreover, 5' nearest neighbor preference is determined by residues in close vicinity. For APOBEC-3A aspartate 131 (D131) and Y132, interact with the Whatson-Crick face of the thymine (T) 5' of the target C. Interactions with a C on that position would not be ideal, as a hydrogen bond with the backbone NH of D131 would not be possible, explaining APOBEC-3A's preference for 5'-TCN codons.[131, 154, 156] What is more, it is believed that Y130 and D131 in APOBEC-3A are the main reason of why a larger purine base would not fit in that position.[154, 157] For APOBEC-3G, which prefers a 5'-CCN codon, interactions with the 5'-C are mediated by corresponding residues D316 and 317.[155, 158] Here, 5'-C forms hydrogen bonds with the sidechain of D316, the backbone NH of D317 and the backbone NH of Q318.[155, 159] What is more, a group reported extensive conformational change of D317 in order to fit in a non-preferred adenine in the 5'-position.[154, 158] Mutational analyses of that residues revealed that substituting the aspartate with tryptophane, as it is found for APOBEC-1 (W121), can increase affinity to a T 5' of the target C even more than a Y, the native residue of APOBEC-3A in that position.[156] Yet, reported 5' nearest neighbor preference of APOBEC-1 is an A.[129]

Interactions with the immediately 3' located base are mostly π-stacking based and are thus not base-specific, which explains the rather lax 5' nearest neighbor preference of APOBECs.[154]

In dual deaminase domain APOBECs such as APOBEC-3G, the N-terminal deaminase domain, despite having no catalytic activity, was shown to be crucial for editing. It can also bind to ssDNA and takes a supporting role in successful deamination.[51, 52, 153] Moreover, binding to RNA was reported to constitute a regulatory mechanism for APOBEC-3G, as deaminase activity is reduced in presence of RNA. This occurs by either RNA molecules competing with ssDNA for binding to the C-terminal deaminase domain or allosterically by binding to the N-terminal deaminase domain, which in turn leads to inhibition of C-terminal deaminase activity.[160]

Dimerization was also described for the APOBECs. In fact, multimerization is a common feature of cytidine deaminases including deaminases of monomeric cytidines.[161] Even though APOBEC-1 was first described as an RNA editing enzyme, it has very low RNA-binding capacity.[162] Consequently, it requires co-factors for target recognition and efficient RNA editing. Before discovery of RBM47, the consensus of APOBEC-1-mediated ApoB mRNA editing had been a model, in which an APOBEC-1 dimer is positioned for deamination by ACF. Responsible residues for APOBEC-1 dimerization are located at the C-terminus, but the extent of implication of dimerization on editing is not entirely elucidated.[85, 86] Most intriguingly, DNA editing of APOBEC-1 does not require any co-factors.[102, 103]

Within the APOBEC-3 family some members multimerize, while others remain monomers in their active form.[51, 52] For oligomerization of APOBEC-3G, two mechanisms were described. Firstly, protein-protein interactions between two C-terminal or two N-terminal deaminase domains form a homodimer, which in turn forms an editing active homotetramer on an ssDNA molecule.[163, 164] In contrast, RNA-binding dependent oligomerization was described as a regulatory mechanism leading to deamination inactive multimeric complex.[165] Most interestingly, APOBEC-3G tends to bind structured RNAs of cellular or viral origin and thus is packaged in budding HIV virions for activation upon subsequent infection.[163, 166] In contrast, APOBEC-3A remains monomeric while editing ssDNA but a dimeric state was also observed.[163, 167] If activity on RNA substrates is monomeric or multimeric has not been described, yet.

Despite similarities in the catalytic center and subsequent deamination mechanism, structural differences demonstrate the divergence of ADARs and APOBECs. Both rely on

RNA secondary structures, but the function of that structure is different. While accurate editing by ADARs require secondary structures, since ADARs arbitrarily edit A within a perfect double helix, APOBECs catalytic center lies within a U-shaped groove, in which the hairpin structure can easily fit. It is however remarkable that seemingly single residues within ADAR's deaminase domain are determinants for why adenosines but not cytidines are edited.

1.2. Application of targeted RNA editing

Despite single nucleotide alterations A-to-I and C-to-U editing have a great impact on various cellular processes. In fact, altered editing levels are connected to various cancers, disease of the nervous systems impacting its function or development (e.g. Aicardi-Goutières syndrome, temporal lobe epilepsy, etc.).[24, 54, 56] Consequently, given this significance, the enormous potential of controlling and employment of this mechanism was seen very early. In fact, in 1995 Woolf et al. for the first time managed to perform targeted RNA editing with endogenous ADARs of *in-vitro* synthesized reporter mRNA hybridized to a complementary oligo first in *Xenopus* oocyte extract and later microinjected into *Xenopus* oocytes.[168] Since then multiple techniques have been developed for site-directed RNA editing. While approaches themselves differed from one another, they can be classified in two main categories: editing by ADAR fusion proteins and editing by endogenously expressed ADARs. For the former, the editase is an artificial fusion protein of the deaminase domain of one of the ADARs and a protein moiety (of eucaryotic, procaryotic, or viral origin) that mediates binding to the target site. More accurately, this protein moiety binds to a guideRNA, which not only provides the necessary structures (e.g. double-strand, positioning of target base) for editing but also contains an element for interaction with the fusion protein.

The latter of the two categories relies on the ubiquity of endogenous expression of ADARs in tissues.[24, 25] Therefore, the guideRNA is the only necessary component to be administered. The antisense oligonucleotide forms the double strand and A:C mismatch. In addition, it can also have a recruiting moiety to steer the ADARs to the target site.

1.2.1. SNAP-fused Editors

A pioneering approach for targeted RNA editing relies on SNAP-tag for mediating recruitment of the deaminase domains of ADAR 1 and 2 (Figure 3, 2nd panel, 4th schematic).

SNAP-tag was generated by extensively mutating endogenous human O^6-alkylguanine-DNA alkyltransferase (hAGT), an enzyme responsible for repair of toxin-induced alkylation of DNA.[169-173] By transfer of the alkyl to one of its cysteines, hAGT is inactivated and subsequently degraded.[172, 174] Therefore, it is referred to as a self-labeling or suicide protein.[172] The introduced mutations resulted in a high affinity to O^6-benzylguanine (BG) as a substrate.[169-172] Further development of SNAP-tag even resulted in an enzyme named CLIP-tag that preferred O^2-benzylcytosine (BC) over BG.[175] Similar to this, a self-labeling enzyme of bacterial origin could also be generated. The dehalogenase of *Rhodococcus rhodochrous* was mutated to covalently bind to chloroalkane resulting in HaloTag.[176]

A remarkable feature of SNAP-tag constituted the high tolerance to residues bound to the benzyl e.g., fluorescent tags or biotin.[177-179] For targeted RNA editing this was exploited by equipping a guideRNA with a BG-moiety *in vitro* to steer the ADAR deaminase domain fused to SNAP-tag to target sites on mRNAs first *in vitro* and later in cell culture.[180-182] For the latter, cells already expressing SNAP-ADAR1/2 by pDNA transfection[181, 183, 184] or doxycycline-induced[182, 184] were only transfected with the BG-tagged guideRNA. Conjugation to SNAP-ADAR and subsequent translocation to the site of interest occurred within the cells. Besides recruitment of the editase, the guideRNA also created the necessary double-strand with a C opposite of the target A to provide the essential mismatch for editing (from here on out referred to as mismatch C).[180] Like this, deaminase domains of both catalytically active ADARs and their hyperactive EQ mutants could very successfully be applied for editing of multiple endogenously expressed transcripts in both their UTR and open reading frames (ORF).[182] Targeting of all possible 5'-NAN codons within the GAPDH transcript revealed an extensive codon scope of both SNAP-ADAR variants. In fact, all codons, except for 5'-GAN codons, were highly edited by at least one of the SNAP-ADARs.[181, 182] With the deaminase domains of the ADARs showing high tolerance towards modifications, the guideRNA could be extensively modified,[181] so much so, that by-stander editing of 5'-NAA or 5'-AAN codons could be suppressed or even completely abolished.[182]

Further developments of the system include masking of the BG-moiety with a light-sensitive protection group to allow covalent binding to SNAP-tag in a lightdose-dependent manner, even in an invertebrate animal model.[183, 184] Also, Stroppel et al. created another approach with the help of the plant-derived gibberellic acid. Here, dimerization of beforehand split SNAP-tag and ADAR deaminase domain could be triggered in a dose-

dependent manner to chemically induce targeted A-to-I editing even in disease-associated mutation (Figure 3, 3rd panel, 4th schematic).[185]

1.2.2. Cas13-based approaches

With the discovery of the clustered regularly interspaced short palindromic repeats (CRISPR) and the CRISPR-associated nuclease 9 (Cas9), targeted manipulation of eucaryotic genomes was severely facilitated ranging from targeted A-to-I or C-to-U DNA editing to gene correction of disease-associated mutations.[186-188] The newest members of the CRISPR world constituted the mRNA targeting Cas enzymes.[189]

By fusing nuclease-dead Cas13b of *Prevotella sp. P5-125* (dPspCas13b) to the deaminase domain of ADAR2Q, Cox and colleagues created the first version of RNA Editing for Programmable A to I Replacement (REPAIRv1).[190] By addition of a hair-pin structure (direct repeat) 3' of the antisense part of the guideRNA, REPAIRv1 is tethered to it and directed to the target site (Figure 3, 3rd panel, 1st schematic). Different from the SNAP-ADAR approach, REPAIR was limited to the ADAR2Q deaminase domain, as other fusions exhibited more severe dependency on long guideRNAs (i.e., ≥70 nt). To further increase specificity to the hyperactive ADAR2Q deaminase domain, T375, which contacts the orphan base flipped out of the duplex was mutated to a glycine (REPAIRv2).[142, 190]

This rational mutagenesis proved effective, as it could increase specificity of the enzyme. As mentioned above, deaminase domains of ADARs and cytidine deaminases show homologous functional residues with a similar reaction mechanism and single residues presumably responsible for not accommodating a cytosine in the catalytic center (section 1.1.5). Consequently, Abudayyeh et al. introduced three mutations and managed to evolve the deaminase domain of ADAR2Q to perform cytidine deamination. Tethering to the guideRNA was mediated by a catalytically dead Cas13b ortholog of *Riemerella anatipestifer* (dRanCas13b). The first mutated residue was valine 351 (to glycine), a residue that directly contacts the target A.[142] The next was, serine 486 (to alanine), which was described to contact the amine of the guanine in a 5'- NAG context.[142] T375 was mutated to serine, a mutation that was shown to decrease both off-target and on-target A-to-I editing for REPAIR.[190] In addition, the side chain of T375 was proposed to clash with the ribose moiety of pyrimidines and thus prevent ADARs from editing them.[141]

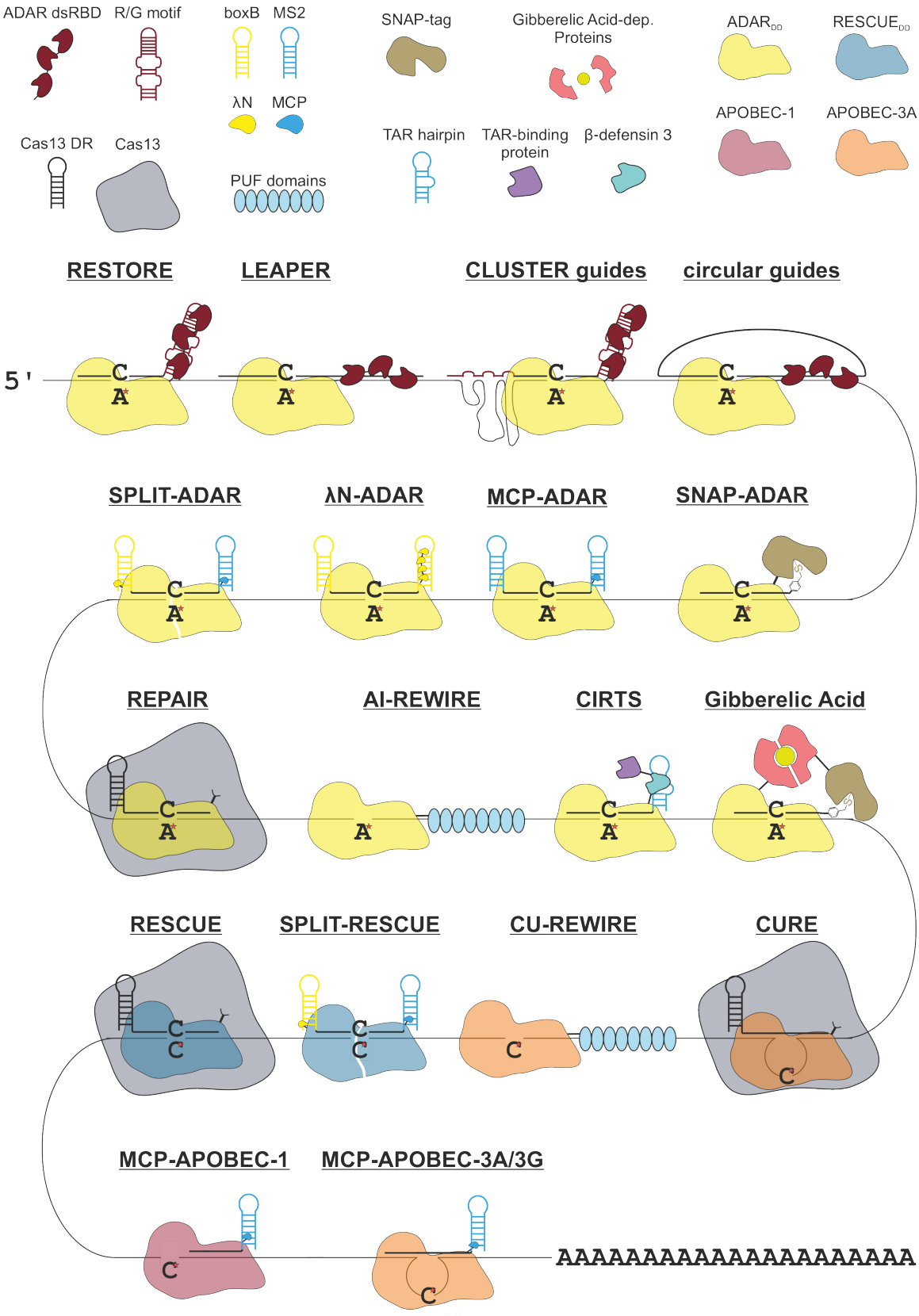


Figure 3: ***Tools for targeted RNA-editing.*** *Schematic depiction of developed tools for targeted A-to-I and C-to-U editing. While some tools recruit endogenously expressed ADARs by using chemically modified or encodable guideRNAs, others rely on a fusion of ADAR's deaminase domain to a protein that interacts with significant moieties (secondary structure, chemical compounds, etc.) within guideRNAs. For targeted C-to-U editing, only fusion proteins are reported that utilize either endogenous cytidine deaminases such as APOBECs or a mutated ADAR2 deaminase domain. Information for schematic visualization taken and adapted from* [180-185, 190-209]

13 rounds of directed evolution then established a highly active version of RNA Editing for Specific C to U Exchange (RESCUE). However, to decrease substantial transcriptome-wide off-target deamination of both, adenine, and cytosine, Abudayyeh and colleagues mutated serine 375 (natively threonine 375) to alanine. Unfortunately, this RESCUE-S did not only exhibit reduced off-target but also on-target editing. Nevertheless, the nature of the system allowed for simultaneous A-to-I and C-to-U editing by application of a pre-guideRNA construct that would be processed into two guideRNAs by the Cas13b ortholog. Despite the weaknesses of the system, it was impressively shown that directed mutagenesis of the ADAR2 deaminase domain could yield an effective cytidine deaminase (Figure 3, 4th panel, 1st schematic).[191]

Recent advances for Cas13-based platforms include the discovery of a compact Cas13 ortholog of a size of 408 amino acids (in comparison RanCas13 1096 amino acids[191]). REPAIR based on this Cas13 ortholog managed to elicit editing when employed as an adeno associated virus (AAV). However, in RESCUE context this Cas13 ortholog poorly yielded editing on endogenous targets (~ 5 %).[210]

Beside the compact Cas13, dCasRx[211] was used to create REPAIRx and extensively benchmarked it to every fusion-based editing platform except for SNAP-ADARs.[212] In another report, RESCUE deaminase domain fused to dPspCas13b named eRESCUE exhibited strongly elevated on- and off-target editing.[191, 213]

Besides RESCUE, further Cas13 fusion proteins were reported for targeted cytidine deamination. C>U RNA Editing (CURE) utilizes APOBEC3A with a tyrosine 132 to aspartate mutation for enhanced activity as the catalytically active moiety and tethers to the guideRNA either by dPspCas13b (CURE-N, -C1/2) or by dCasRx (CURE-X).[195, 211] Inspired by the secondary structure of one of the endogenous targets of APOBEC-3A, R46 in SDHB (Figure 1B), the antisense part of the guideRNA forces the target mRNA to form a loop containing the target cytidine (Figure 3 4th panel, 4th schematic). CURE editases elicited high-level RNA editing but only within a strongly limited codon scope of 5'-UCN as according to APOBEC-3A's natural nearest neighbor preference.[62, 131]

1.2.3. Bacteriophage-based systems

Further approaches with entirely encodable components were based on bacteriophage mechanisms relying on an interplay between a protein and RNA secondary structures. While the protein portion was fused to the deaminase domain of the editing enzyme for

mediating translocation to target sites, the conserved sequence for the hairpin formation was added to the guideRNA, for recruitment of the editase.

Bacteriophage λ regulates its gene expression within the host by an interaction of the N-protein with the bacterial RNA polymerase and a hairpin structure within the operon called boxB.[202, 203, 207, 214-216] As only 22 amino acids of the λN-protein are required for binding, they are fused to the deaminase domain of ADAR2 or its hyperactive EQ variant and the boxB motif is fused to the guideRNA for targeted A-to-I editing.[202, 203, 207] Like this a powerful tool was created with fully encodable components that could efficiently edit disease-associated mutations.[202, 203, 207] The small size allowed for packing into AAV vectors and delivery to primary neurons (Figure 3, 2nd panel, 2nd schematic).[207]

Another approach with completely encodable components was based on the bacteriophage R17 coat protein MCP.[192-194, 196, 208, 217] Similar to the λN-peptide, the MCP recognizes the secondary structure MS2, which is fused to the guideRNA.[192-194, 196, 208] While this system initially, exhibited subpar editing on reporter transcripts, it showed potential for improvement.[192, 193, 196, 208] Like for SNAP-tag, both catalytically active ADAR deaminase domains and their corresponding hyperactive variants could be recruited, respectively for considerable editing on both, reporter, and endogenous transcripts even *in-vivo* (Figure 3, 2nd panel, 3rd schematic).[196, 208] In addition, the MCP-system also allowed for targeted C-to-U editing on a reporter transcript by using APOBEC-1 (Figure 3, 5th panel, 1st schematic).[194] What's more, in a very recent report, the MCP-system was used for editing with APOBEC-3A and a mutated APOBEC-3G with a guideRNA design resembling the one employed by the CURE system (Figure 3, 5th panel, 2nd schematic).[199]

Most intriguingly, in a combinatory approach, deaminase domains were split and either part was fused to MCP and λN-peptide to be assembled by a guideRNA bearing both boxB, and MS2 hairpins. Like this transcriptome-wide off-target by ADAR (Figure 3, 2nd panel, 1st schematic) and RESCUE (Figure 3, 4th panel, 2nd schematic) deaminase domains were severely reduced.[197]

1.2.4. Fusions with human and other parts

CRISPR-Cas – Inspired RNA Targeting System (CIRTS) constitutes a multivalent system with great potential, as it consists of multiple protein domains performing different task fused to each other (Figure 3, 3rd panel, 3rd schematic). The rationale of this approach was to provide an alternative to the CRISPR/Cas proteins, which are relatively large, and a

considerable number of individuals were reported to already have antibodies against them.[218] Therefore, functions such as nuclease-activity, or any other effector-dependent activity for that matter, or interaction with RNA, which are all tasks performed by Cas13 had to be substituted. Unfortunately, there is no known human enzyme to perform all those tasks. Subsequently, it was sought to take domains of known proteins and combine them into one multidomain construct. Like this, it was possible to create a modular system, whose parts could be exchanged to perform desired tasks, ranging from knock-down of target mRNAs, to enhancing their translation and performing targeted A-to-I editing.[205]
Another exciting approach constituted RNA-editing with individual RNA-binding enzyme (REWIRE).[219] This single molecule system does not require a guideRNA, as target recognition is mediated by RNA binding domains of the Pumilio and FBF (PUF) protein family (Figure 3, 3rd panel, 2nd schematic & 4th panel, 3rd schematic). PUF proteins are found in all eukaryotes, directly bind mRNAs are thus involved in their post-transcriptional regulation (reviewed here[220]). The binding domains consist of eight alpha-helix repeats, every one of which recognizes one RNA base, respectively.[220] Consequently, those repeats can be arranged in any way to fit any RNA sequence for its subsequent manipulation.[219]

1.2.5. Endogenous ADAR

A therapeutically interesting approach constitutes utility of endogenous levels of ADARs for targeted A-to-I RNA editing. This way, the only therapeutic to be administered is the oligonucleotide. In 2017 Wettengel et al. fused a hairpin structure of the endogenous R/G target on GluR-B (here referred to as R/G-motif) to the 5'-end of an 18 nt long antisense part of the guideRNA.[125, 209] Like this, it was demonstrated that over-expressed ADAR2 was successfully recruited for editing on even disease-associated mutations by plasmid-borne guideRNAs.[209] On this basis, Merkle et al. developed recruiting endogenous ADAR to specific transcripts for oligonucleotide-mediated RNA editing (RESTORE).[201] Here, chemically modified guideRNAs bearing modified versions of the R/G-motif and a longer (40 nt) antisense part managed to recruit all ADAR isoforms in both, various cell lines, and primary cells even functionally restoring a disease-relevant mutation (Figure 3, 1st panel, 1st schematic).[201]
Another system relying on leveraging endogenous ADAR for programmable editing of RNA (LEAPER) resembles the approach employed by Woolf et al. in 1995, by which a guideRNA forming an unstructured dsRNA with the target transcript could elicit editing

by endogenously expressed ADAR (Figure 3, 1st panel, 2nd schematic).[168, 204] LEAPERs guideRNAs are long (over 71-151 nt for editing on endogenous transcripts), chemically unmodified, and applied as plasmid DNA under U6-promoter triggered expression. As ADARs are known to elicit editing indiscriminately in RNA duplexes, the authors here also observed by-stander editing sites, which they largely contained by base-pairing adenines in questions with guanines.[168, 204]

Another approach, which is part of this thesis, is also based upon encoded guideRNAs. Different from LEAPER, theses CLUSTER guideRNAs consist of an antisense domain to the target site with an R/G-motif for ADAR recruitment (Figure 3, 1st panel, 3rd schematic). In addition, these guideRNAs contain antisense domains that bind at various sites throughout the targeted transcript called recruitment sequences (RS). The RS are designed so they bind to sequences absent of As in ADAR-preferred sequence contexts, as to avoid by-stander editing, as observed for long LEAPER guides. These CLUSTER guideRNAs elicited efficient and accurate editing of both over-expressed, and endogenous transcripts even in disease-relevant contexts. Ultimately, delivery via hydrodynamic tail vain injection allowed for targeted A-to-I editing of overexpressed reporter transcript in mice by recruitment of endogenous murine ADAR by CLUSTER guideRNAs.[206]

The most recent development for recruitment of endogenous ADARs makes use of autocatalytic cleavage of specific RNA sequences (i.e. Twister sequences) flanking the target-specific antisense part of the guideRNA. Upon cleavage of the Twister sequences, the ubiquitously expressed RNA ligase RtcB ligates the ends of the guideRNA thus forming circular guideRNAs (Figure 3, 1st panel, 4th schematic). These cicular guideRNAs elicited solid A-to-I RNA editing even in a disease-relevant mouse model when delivered as AAV.[198, 200]

1.3. Chemical modifications

Chemical modifications have dramatically increased potency of various oligonucleotide-dependent applications by providing protection from nucleases or increasing binding affinity to target mRNAs. Besides stability, these chemical modifications have allowed to abstain from utility of chemical carriers commonly used for delivery to cells and tissues and thus enabled for free or gymnotic uptake.[221-225] Here, modifications can dictate interaction with specific or unspecific receptors for productive uptake, that is, the oligonucleotides arrive fully functional at the site of action. However, despite the positive

effects, a simplistic profusion of chemical modifications can be detrimental to the application. Therefore, adjustment of oligonucleotide design and chemical patterns to the specific application to maintain activity while retaining all positive effect is imperative. As the landscape of chemical modifications is vast, this chapter will be focused on the most commonly applied modifications, but the reader is referred to these reviews for further reading.[222-225]

1.3.1. Modification of bases

Chemical modifications of nucleobases are very common within cells serving multiple tasks, some of which are masking from pattern recognition receptors, skipping of stop codons, or recruitment of effector proteins. For applications that utilize mRNA for transgenic expression in a gene supplementation approach or vaccinations, applied chemical modifications can strongly reduce innate immune response towards the mRNA itself. In addition, chemically modified bases can also be employed in RNase H-dependent antisense oligonucleotides (ASO).[224, 226-228]

For A-to-I RNA editing specific base modification can have a severe impact on editing efficiency. To re-iterate, the target A is flipped out of the dsRNA substrate for deamination. For stabilization, E488 (ADAR2)/E1008 (ADAR1) invade the double-strand from the minor grove side and interacts with the orphan cytosine left by adenine flipping out. Structural analyses have shown that G:C base pairing immediately upstream of the target A impairs strand invasion by E488/E1008 due to guanine's amino group clashing with the flanking glycine (section 1.1.5, Figure 2C). Here, introduction of a C:I base pair can strongly increase editing, as hypoxanthine (the base of inosine) lacks guanine's obstructive amino group (Figure 4A).[142]

Moreover, the hyperactivity of the E488Q/E1008Q mutants was shown to be based upon ideal H-bonding of the Q mutant with the orphan cytosine at physiological pH. By utilizing a cytidine analog at the orphan base position, reported editing levels with wildtype ADAR were higher than with the hyperactive EQ mutant. This Benner's base can form an H-bond with E488/E1008 independent of pH and thus strongly increases editing (Figure 4A).[229]

1.3.2. Ribose

Regarding the ribose moiety of the RNA, the 2' OH allows for extensive possibilities for modifications (Figure 4B). The most basic being just a hydrogen forming a naturally

occurring deoxyribose. Applications such as RNAse H-dependent ASO rely on that in order to form an RNA:DNA heteroduplex to trigger RNase H-dependent degradation of target RNAs.[230]

Another naturally occurring and commonly used modification is 2'O-methyl (2'OMe). This modification provides strong protection from nuclease degradation and can therefore strongly increase half-life of oligos in serum and other harsh medium. In addition, this modification promotes the ribose to switch into the 3' endo pucker, by which binding affinity to target RNAs is increased.[231-233] Unlike RNA-induced Silencing Complex (RISC), RNase H and ADARs are very intolerant towards this modification at specific regions of the RNA:olignucleotide duplex.[168, 223, 224] In fact, A-to-I editing with ADARs can be extinguished when utilized oligonucleotides are entirely 2'OMe modified. More precisely, when the nucleotide opposite of the edited A is 2'OMe modified, editing is impaired.[168] For applied A-to-I editing, both the mismatch C, and the flanking nucleotides are left unmodified on the ASO, as to not interfere with editing.[181, 182, 201] However, applying modification close to the mismatch C can reduce by-stander editing within A-rich contexts.[182] To increase stability of RNase H ASOs, gapmers were developed, in which a stretch (or gap) of deoxynucleotides of sufficient length was flanked by other 2'-modifications that promote stability from nucleases. For the first developed gapmers a modification pattern was established in which nucleotides flanking the critical deoxynucleotide center of the gapmers bore 2'OMe.[230, 234] Similarly, the 2'-methoxyethyl (2'MOE) modification is employed at the wings of gapmers, as it also increases ASO affinity and provides excellent protection from degradation.[233, 235] In addition, this modification is employed in steric blocker ASOs. These ASOs are fully modified and do not elicit RNase H-dependent degradation of target RNAs. Instead, applied ASOs hybridize with the target mRNA eliciting various effects, e.g. exon skipping.[224, 236]

Another commonly utilized modification is a fluorine at the 2'-position (2'F). It strongly increases binding affinity but protection towards nucleases is not as strong as with the other modifications when applied alone.[237, 238] 2'F is employed in aptamers and siRNAs. In fact, for the latter a pattern of alternating 2'F and 2'OMe was established.[224, 239, 240] Same as for 2'OMe, 2'F can help reduce by-stander editing sites for targeted A-to-I editing of A-rich codons.[182]

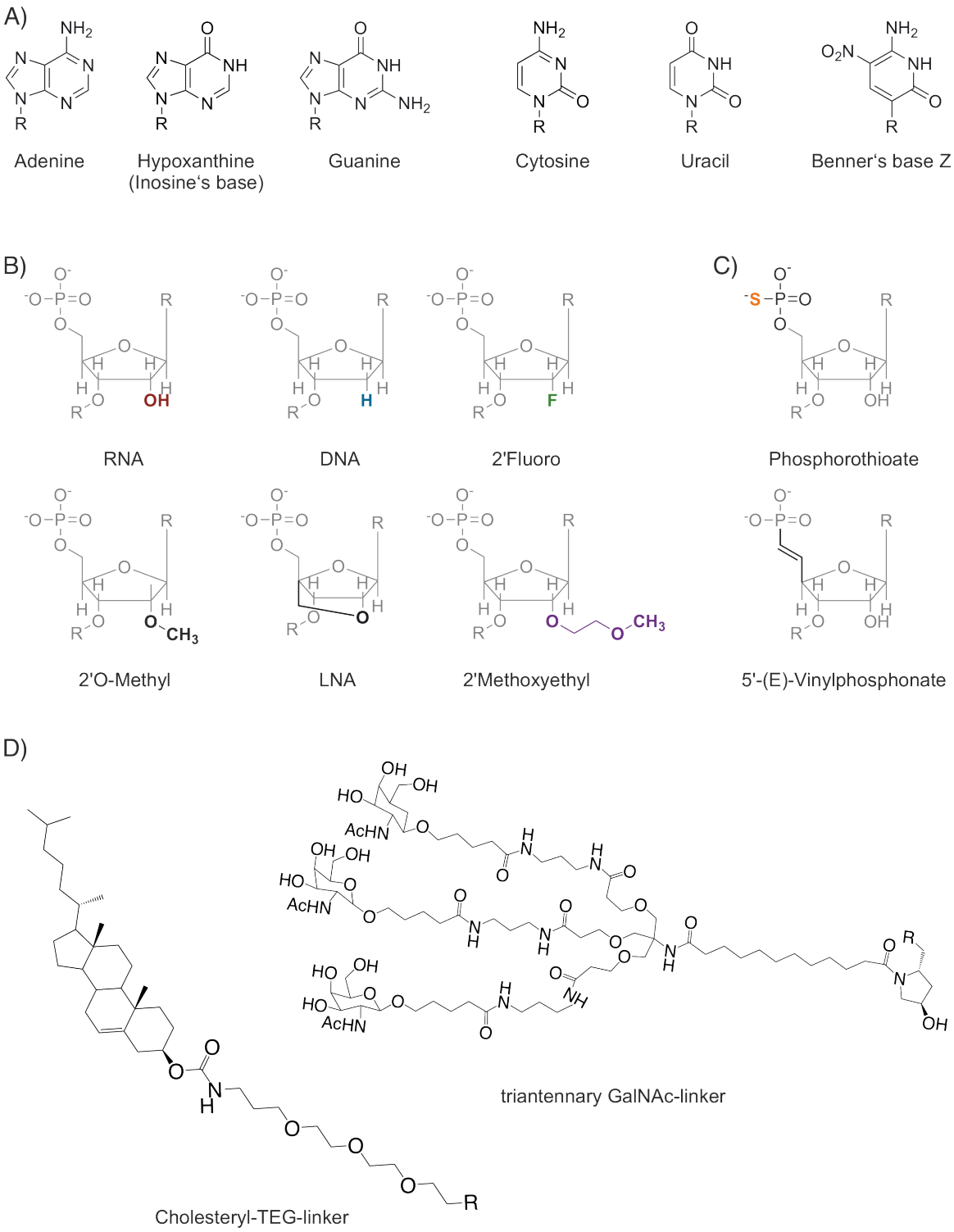


Figure 4: ***Common modifications employed in oligonucleotide applications.*** ***A)*** *Nucleobases found in RNAs. Deamination of Adenine results in Hypoxanthine, which possess similar biological features as Guanine. In addition, Hypoxanthine lacks the NH2-group of Guanine, which interferes with deamination by ADARs. Deamination of Cytosine results in Uracil. Benner's base Z is a Cytosine ortholog that forms ideal hydrogen bonds with the strand invading ADAR's E488/E1008 independent from pH. This way with wildtype ADARs editing levels of hyperactive Q-variants is achieved.* ***B)*** *Commonly used ribose modifications. Different 2'-substituents are depicted in different colors.* ***C)*** *Modifications of the Phosphate backbone. Substituting one of the oxygens not involved in polymerization by sulfur (upper structure in red) leads to phosphorothioate a modification various beneficial features for the oligonucleotides and subsequently employed in many oligonucleotide applications. 5'-(E)-Vinylphosphonate is a modification known used in siRNAs to provide protection from phosphatase while preserving functionality and thus strongly increased in-vivo siRNA efficacy.* ***D)*** *Different linkers for delivery to cells/tissues. Cholesterol (left) can mediate uptake via scavenger receptors and can interact with lipoproteins. Triantennary N-Acetylgalactosamin (GalNAc, right) is commonly used for targeting*

hepatic tissue for asialoglycoprotein receptor (ASGPR)-mediated uptake. Information for visualization taken and adapted from [142, 223, 224, 229].

Lastly, by covalently binding the 2' and 4' carbon, the ribose is constraint and locked into the 3'-endo pucker and provides the strongest thermostability of duplexes.[241] One variation of this bridged nucleotide modification constitutes the locked nucleic acids (LNA).[242, 243] This modification increases not only binding affinity to complementary sequences but also provides nuclease protection. Consequently, it is found in various ASO applications.[222, 224, 241]

1.3.3. Phosphate

The phosphate backbone of the DNA constitutes a vital point for oligonucleotide applications. By substituting one of the oxygens, not involved in nucleotide polymerization, with sulfur, not only stability but also chemical features of the oligo can be altered (Figure 4C). Oligonucleotides containing phosphorothioate (PS) show a reduced binding affinity, increased nuclease resistance, and can interact with cellular and plasma proteins (e.g. albumin), the latter of which reduces clearance of the oligos via kidneys.[244] In addition, PS-modified ASO can be taken up by scavenger receptors, thus allowing for delivery to the cell unassisted by cationic lipids in a free and productive uptake.[221, 245] Most intriguingly, this modification creates a chiral center around each phosphorus of the oligonucleotide. Consequently, oligonucleotides with a PS backbone are a racemate with n^2 different possibilities with n being the number of nucleotides within the oligonucleotide. Interestingly, these stereochemical configurations named Rp and Sp can impart different characteristics to ASOs. While an Rp stereopure ASO exhibits slightly increased binding affinity to target sequence, an Sp stereopure ASO is more stable in harsh medium and is more lipophilic.[246, 247] However, despite the overall positive effects of the Sp stereochemistry, a position-dependent employment of either Sp or Rp is important for efficiency of RNase H ASOs.[247, 248]

Like for other modifications, PS modifications are also not tolerated by all applications. While RNAse H dependent gapmers make regular use of PS, siRNA lose efficacy when PS modification is too extensive, which is why they are employed close to the ends of the strands.[249-251] Similarly, PS modification at sites remote from the center of the oligonucleotide are well accepted by ADAR deaminase domains.[181, 182, 201]

1.3.4. Termini

While PS or ribose modifications impart strong resilience towards nuclease degradation, the termini of the ASOs open up entirely new possibilities. The double-stranded structure of siRNAs is here an advantage, as they have four possible termini that can be modified.[223] Employed modifications can, on the one hand, aid for delivery to cells in both, a specific and unspecific manner, or they are of requisite importance for the function of the system. For siRNAs, guide strand 5' phosphorylation is indispensable for RISC assembly but shows instability in serum. Therefore, 5'-(E)-vinylphosphonate can be used to mimic phosphorylation while promoting stability to phosphatases and preserve functionality (Figure 4C).[252, 253] The same effect was also shown for PS and other modifications at that end.[254] For SNAP-ADARs, the SNAP-tag substrate is conjugated to the 5'-end of the guideRNA via an amino linker. Only then can the SNAP-ADAR be steered to the target site for A-to-I editing.[181, 182]

On the other hand, the termini can be equipped with a compound that can mediate targeted uptake in organs or tissues (Figure 4D). A triple N-Acetylgalactosamin (GalNAc) is a broadly used compound for targeted delivery to hepatic tissue. GalNAc interacts with the asialoglycoprotein receptor (ASGPR), which shows high expression in hepatocytes, and is subsequently internalized via endocytosis. In an acidic milieu such as endosomes, GalNAc then disassociates from ASGPR and is enzymatically cleaved. These features, make GalNAc an ideal compound for ASO delivery to hepatic tissue and is therefore broadly used in ASO applications from cell culture to animal models and even in the clinic.[224, 255-257]

Moreover, employment of lipophilic molecules such as cholesterol can promote cellular uptake either via interaction with scavenger receptors or by association with lipoproteins and subsequent uptake via their respective routes. The latter can even determine tissue distribution of ASOs. Incubation of cholesterol-tagged siRNAs with LDL prior to application leads to exclusive hepatic uptake, while HDL mediated uptake in multiple other organs besides the liver.[258-261]

1.4. Aim of the thesis

In the light of the great potential of targeted RNA editing, multiple tools with various up- and downsides emerged in recent years. While some rely on completely encodable components, the SNAP-ADAR system relies on chemically modified guideRNAs. Like this it was not only possible to elicit very high editing yields on a broad spectrum of different codons across multiple endogenous transcripts with both ADAR deaminase domain in addition to their respective hyperactive variants but also fine-tuning of the guideRNA's chemistry allowed for control of unwanted by-stander editing.[180-182] In this work we seek to expand this highly effective system in multiple ways.

Firstly, given the tolerance of ADAR's deaminase domain towards chemical modifications, we seek to further explore their employment. Other ASO applications have shown that those chemical modifications can have very positive effects by increasing both stability and affinity of employed oligonucleotides. Here, we seek to apply known chemical modifications to increase stability of guideRNAs in hostile environment, to which they are very like exposed on their way to the target site, while simultaneously maintain or even improve editing capacity especially for difficult to edit codons. What is more, we seek to develop a modification pattern that allows for carrier-free application of guideRNAs.

Secondly, while the SNAP-tag constitutes an excellent self-labeling enzyme for targeted A-to-I RNA editing, other enzymes with similar functions could allow us to add another layer of versatility. Consequently, we seek to expand the SNAP-tag system with another self-labeling enzyme that could allow for concurrent and orthogonal employment of various effectors. Like this, simultaneous A-to-I and C-to-U editing could be enabled and thus expand the scope of effects elicited on cellular physiology.

Lastly, unlike A-to-I editing, control of C-to-U editing is a rather new accomplishment with serious limitations ranging from low efficiency to severely restricted codon scope. Here, we seek to provide a powerful alternative using SNAP-tag and chemically modified guideRNAs that do not suffer from those short comings to effectively advance the field of targeted RNA base editing.

2. Results and Discussion

2.1. Recruitment of ADAR with CLUSTER guideRNAs

The following section discusses findings in the development of a fully encodable guideRNA design for recruitment of endogenous ADAR for targeted A-to-I editing. The main part of the results discussed in this section are part of the doctorate thesis of Philipp Reautschnig, whom I assisted in this project. As this does not constitute the main part of my doctorate thesis, I will briefly summarize and discuss key finding in this project. My personal contribution to this project can be found in the Publications section under Publication 3.

In recent years a plethora of different tools has been developed for targeted A-to-I editing. While some rely on over-expressed artificial fusion proteins, others recruit endogenously expressed ADARs. Here, applied guideRNAs are either plasmid encoded, or they are chemically modified. While the latter shows formidable efficacy in multiple various cell lines and primary cells[201], the former provides the opportunity to encode guideRNAs in viral vectors and thus facilitating delivery to hard-to-reach tissues. Consequently, in 2017 Wettengel et al reported an encodable guideRNA bearing the R/G motif (section 1.2.5) 5' of a specificity sequence antisense to the target RNA for recruitment of ADARs for targeted A-to-I editing.[209] In 2019 Qu and colleagues then reported recruitment of endogenous ADARs by an unstructured and encodable guideRNA slightly reminiscent of the approach of Woolf and colleagues from 1995.[168, 204] To elicit editing on endogenous transcripts, these LEAPER guideRNAs were over 100 nucleotides long and were thus accompanied by by-stander editing events of As within the guideRNA:mRNA duplex. In stark contrast, CLUSTER guideRNAs consist of a 20-nucleotide long specificity domain complementary to the target site with a 5' located R/G motif. In addition, at its 3'-end the guideRNAs possess a cluster of single-stranded recruitment sequences (RS) varying in length, which show complementarity to sequences widely distributed throughout the target mRNA. Like this, CLUSTER guideRNAs strongly bind to the target mRNA, while having a high degree of sequence flexibility (Manuscript 3, Figure 1 & 2). RS are determined by an algorithm developed for this purpose, which designs RS based on sequences within the target mRNA lacking adenosine substrates in preferred nearest neighbor contexts (Manuscript 3, Supplementary Figure S1). The CLUSTER guideRNAs elicited solid editing on both, endogenous transcripts and over-expressed transcripts bearing disease-associated

mutations up to the point of considerable restoration of protein function (Manuscript 3, Figure 3). What is more, while editing levels were comparable to or even higher than those obtained by LEAPER guideRNAs, CLUSTER guideRNAs provided a by-stander control beyond comparison (Manuscript 3, Figure 4). In a direct benchmark, CLUSTER guideRNA elicited not just higher on-target but only low levels of by-stander editing if at all. In stark contrast, LEAPER guideRNAs elicited by-stander editing on multiple sites throughout the target transcript (Manuscript 3. Figure 4). In fact, for one transcript over 20 by-stander sites were detected with editing levels of up to 50% (Manuscript 3, Figure 4C). While it was reported that by-stander sites can be quenched by base-pairing with guanosine, in our hands it did not always invoke the desired effect and when by-standers were successfully quenched than on-target editing was also substantially reduced (Manuscript 3, Figure 4C). To assess transcriptome-wide off-target events, we performed transcriptome sequencing and detected only a handful of *de novo* editing site for both CLUSTER and LEAPER guideRNAs (3 vs. 7, Manuscript 3, Figure 4D). However, most intriguingly, this deep sequencing revealed the extent of the by-stander editing on single read level for the target transcript. While about 88% of reads were without by-stander editing for CLUSTER guideRNAs, only 17% of reads for LEAPER guideRNAs exhibited the same purity (Manuscript 3, Figure 4E & F). Consequently, factoring in editing fidelity, CLUSTER guideRNAs greatly outcompeted LEAPER guideRNAs. Finally, we applied CLUSTER guideRNAs for editing of an over-expressed reporter transcript in wildtype C57BL/6 mice by recruitment of endogenous murine ADAR. Plasmid DNA for both, target transcript and guideRNA were co-delivered by hydrodynamic tail vein injection. 72h post-injection, mice were sacrificed, and the liver was extracted. Editing was determined by detection of reconstituted Luciferase signal and by Sanger Sequencing on RNA level. My personal contribution here was euthanasia of the mice, extraction of the liver and subsequent detection of RNA editing yield. We tested two different CLUSTER guideRNA designs and elicited 5-10 % editing on both reporter, and RNA level (Manuscript 3, Figure 5).

In summary, the computer assisted design of CLUSTER guideRNAs demonstrates a great combination of efficacy and accuracy by eliciting high editing by recruitment of endogenous ADAR, while avoiding by-stander sites. The solid editing yields by naked delivery of plasmids encoding for the CLUSTER guideRNAs to murine liver, suggests that employment as a viral vector could elevate editing yields to new heights and unlock the full potential of CLUSTER guideRNAs even beyond hepatic tissue making them therapeutically particularly interesting.

2.2. Targeted A-to-I editing with chemically modified guideRNAs

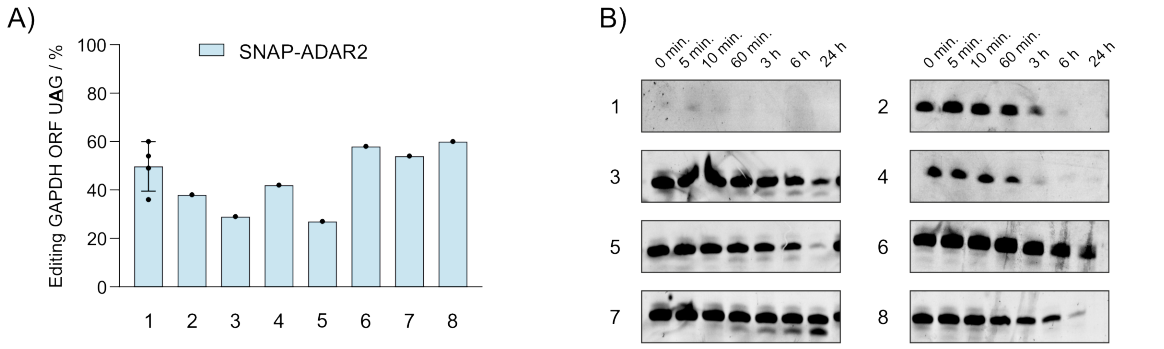
2.2.1. SNAP-ADARs tolerate stability promoting modifications

A pioneering approach for targeted A-to-I editing relies on a fusion protein of the self-labeling enzyme SNAP-tag and the deaminase domain of either of the ADARs.[180] By equipping the guideRNA with SNAP-tag's substrate at its 5'-end, SNAP-ADARs could be directed for efficient A-to-I editing within almost any codon context.[182] There, two major aspects of the guideRNA design played a crucial role in its efficacy. Firstly, for translocating SNAP-editases to the desired site of action, employed guideRNAs formed a double strand at the target site and positioned the target A in a mismatch with a C. Binding to the target site was slightly asymmetric with eight nucleotides 5' of the mismatch C and ten nucleotides 3' of the mismatch C showing complementarity to the target site. In addition, the guideRNA contained a three-nucleotide extension at its 5'-end that did not bind to the target site but provided flexibility for the SNAP-editase and which bore a C6-amino linker (NH), to which the snap-substrate was attached.[180-182] Therefore, this design is here referred to as 5'- (+3)8-C-10. Secondly, it was shown early that using a synthetic oligo bearing exclusively 2'OMe modifications extinguished editing with endogenous ADAR entirely.[168] Consequently, the employed snap-gRNAs were fully 2'OMe-modified except for the mismatch C and its immediately flanking nucleotides (from here on referred to as the "mismatch gap") to provide maximum accuracy. In addition, two phosphorothioate linkages between three 5'-terminal nucleotides and four PS between five 3'-terminal nucleotides were employed.[181, 182] When editing A-rich codons, editing of target A flanking adenosines was observed. This could be suppressed by carefully modifying the mismatch gap. In fact, this effect was obtained by only modifying the sugar moiety of the mismatch C.[182]

Consequently, we sought to further develop guideRNA designs that would provide maximum stability in harsh medium (serum or lysosomes) thus allowing application in animal models, while retaining the high levels of editing capacity reported for SNAP-ADARs.[182]

To this end, we investigated these aspects on a guideRNA targeting an adenosine within a 5'-UAG context within the ORF of glyceraldehyde-3-phosphate dehydrogenase (GAPDH), which was demonstrated to be efficiently edited by all SNAP-ADARs.[182] Stability was determined by incubation of NH-guideRNAs (do not contain snap-substrate) in phosphate

buffered saline (PBS) supplemented with fetal bovine serum (FBS, 9 % final conc.) over a period of 24 h. The guideRNA design remained as published and depicted in Figure 5C.[182] We transfected a HEK 293 Flp-In™ T-REx™ cell line stably expressing SNAP-ADAR2[182] under doxycycline induction with 5 pmol of snap-guideRNA with various modification patterns. Modification patterns were restricted to the mismatch gap and included 2'OMe, 2'F, and combinations of PS modifications. More accurately, only nucleotides flanking the mismatch ribose C (rC) were modified (Figure 5C).



C)

guideRNA	Sequence
1	mU*mA*mU mGmGmU mUmUmU mU **rC rC rA** mGmAmC mGmG*mC* mA*mG*mG
2	mU*mA*mU mGmGmU mUmUmU mU **mC rC mA** mGmAmC mGmG*mC* mA*mG*mG
3	mU*mA*mU mGmGmU mUmUmU mU **fC rC fA** mGmAmC mGmG*mC* mA*mG*mG
4	mU*mA*mU mGmGmU mUmUmU mU **mC* rC mA** mGmAmC mGmG*mC* mA*mG*mG
5	mU*mA*mU mGmGmU mUmUmU mU **fC* rC fA** mGmAmC mGmG*mC* mA*mG*mG
6	mU*mA*mU mGmGmU mUmUmU mU **mC rC*mA** mGmAmC mGmG*mC* mA*mG*mG
7	mU*mA*mU mGmGmU mUmUmU mU **fC rC*fA** mGmAmC mGmG*mC* mA*mG*mG
8	mU*mA*mU mGmGmU mUmUmU mU **rC* rC*rA*** mGmAmC mGmG*mC* mA*mG*mG

rN = RNA
dN = DNA
mN = 2'OMe
fN = 2'F
lN = LNA
N*N = PS

Figure 5: ***Chemical modifications impact editing and stability of guideRNAs. A)*** *FlpIn TReX cells stably expressing SNAP-ADAR2 were transfected with 5 pmol of snap-guideRNAs. Editing was assessed 24 h post transfection. Data shown as mean of ≥1 independent experiments, as indicated by individual points.* ***B)*** *NH-guideRNAs were incubated in PBS containing FBS (9% final concentration) at 37 °C. At indicated timepoints 15 pmol of NH-guideRNA was sampled. Samples were separated on 5M urea 20 % polyacrylamid gel and stained with SYBRGold.* ***C)*** *guideRNA sequence and mofication patterns. Nucleotides in bold represent the mismatch gap of 5'UAG within GAPDH transcript.*

5'- UAG was efficiently edited by SNAP-ADAR2 using a guideRNA with an unmodified mismatch gap but exhibited virtually no stability in 9% FBS (guideRNA #1, Figure 5A & B). Upon addition of 2'OMe and 2'F, editing was reduced with 2'F having a slightly more severe negative effect than 2'OMe (guideRNAs #2 & 3, Figure 5A). In contrast, 2'F provided more stability in 9 % FBS (guideRNA still detectable after 24 h) than 2'OMe (guideRNA almost completely degraded after 3 h, Figure 5B). Adding PS 5' of mismatch rC slightly increased editing for 2'OMe modification pattern and marginally reduced editing for 2'F (guideRNAs #4 & 5, Figure 5A). Strikingly, this modification reduced stability of guideRNA in 9 % FBS for both modification patterns, with 2'F being almost

all degraded by 24 h and 2'OMe by 60 minutes (Figure 5B). In contrast, adding PS immediately 3' of the mismatch rC increased both stability and editing, with the latter exceeding the unmodified guideRNA (#6 & 7, Figure 5A &B). When the gap was fully modified with only PS (#8, Figure 5C) editing and stability were increased as compared to the unmodified guideRNA, but stability was substantially lower than when 2'OMe or 2'F and only one PS modification was employed (Figure 5A & B).

Next, we tested the effect of further modifications on editing with SNAP-ADAR1, its hyperactive isotype SNAP-ADAR1Q and SNAP-ADAR2. Firstly, we introduced a deoxy cytidine as a mismatch C (dC) as the only modification and flanked by 2'OMe modifications, respectively (#9 & 10, Figure 6C). Both slightly increased stability in 9 % FBS but substantially decreased editing for all tested editases (Figure 6A & B).

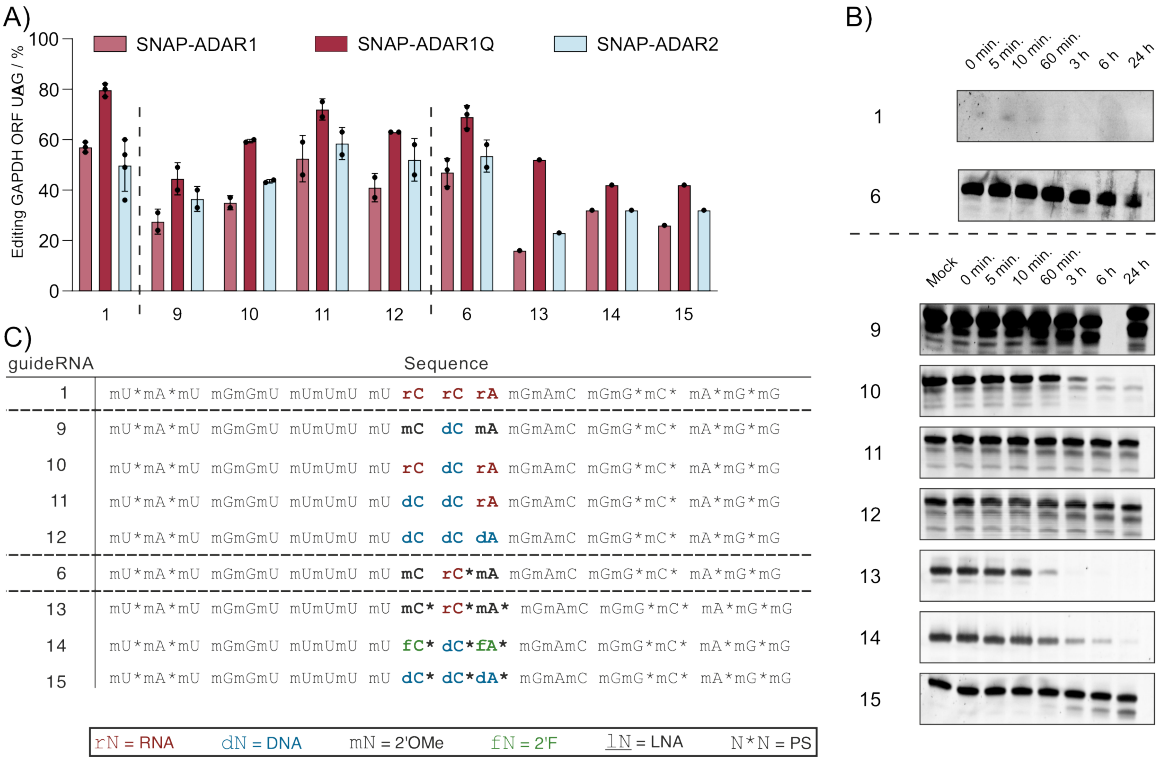
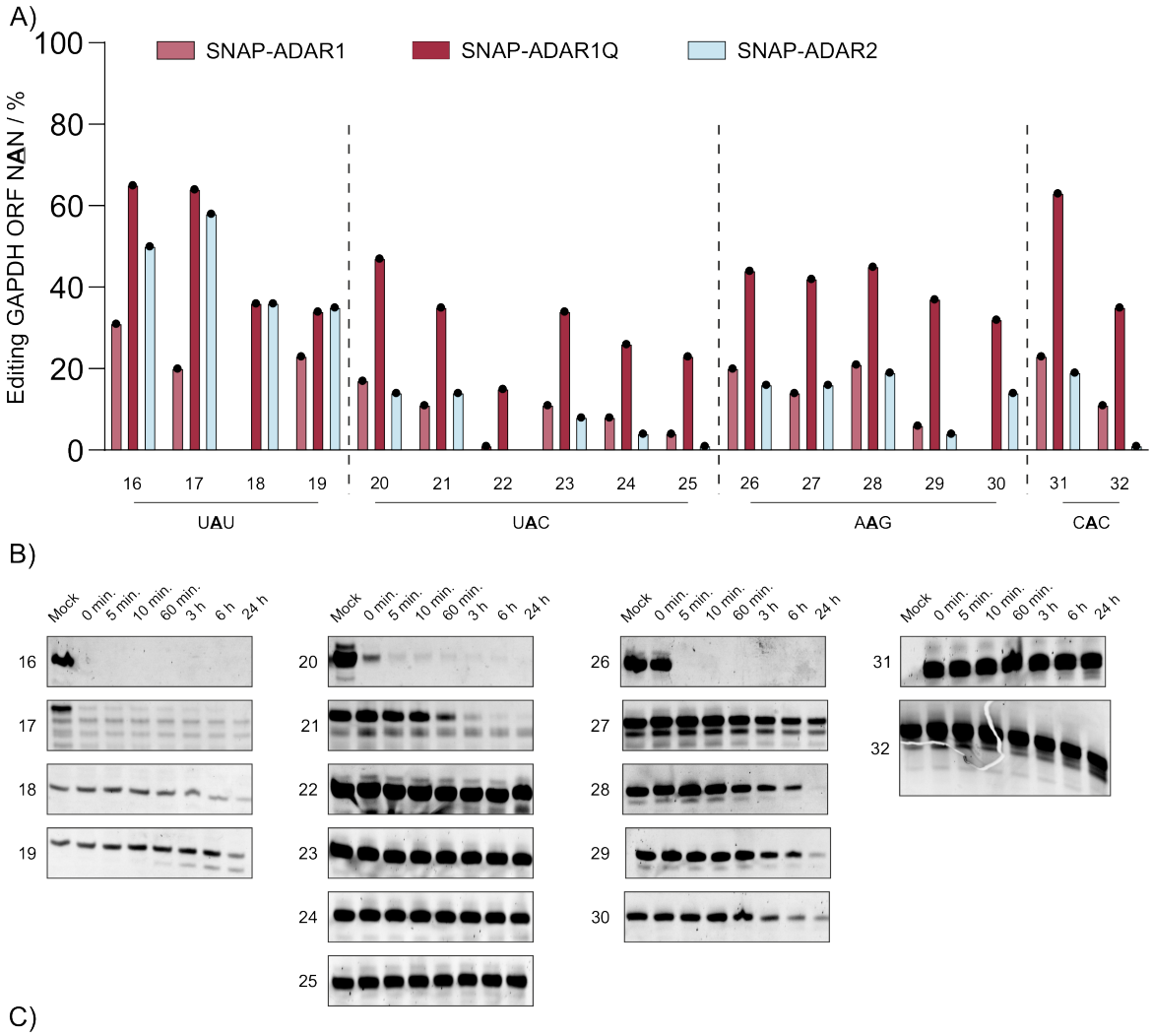


Figure 6: ***Deoxy modification stabilizes guideRNAs.*** ***A)*** *HEK 293 Flp-In™ T-REx™ cells stably expressing SNAP-ADARs variants were transfected with 5 pmol of snap-guideRNAs, respectively. Editing was assessed 24 h post transfection. Data shown as mean of ≥1 independent experiments, as indicated by points.* ***B)*** *guideRNAs were incubated in PBS containing FBS (9% final concentration) at 37 °C. At indicated timepoints 15 pmol of guideRNA was sampled. Samples were separated on 5M urea 20 % polyacrylamid gel and stained with SYBRGold.* ***C)*** *guideRNA sequence and mofication patterns. Nucleotides in bold represent the mismacth gap of 5'UAG within GAPDH transcript.*

Secondly, we introduced a deoxy nucleotide 5' of the mismatch dC, which strongly increased stability to the maximum recorded time-point of 24 h, while only slightly reducing editing (#11, Figure 6). When the mismatch gap was fully deoxy modified

stability remained at maximum but editing decreased slightly more (#12, Figure 6). Fully PS modifying the mismatch gap gave varying degrees of stability and impact on editing (#13-15, Figure 6C). When the mismatch rC was flanked by 2'OMe modified nucleotides, stability plummeted to the point at which only a faint band was detected at 60 minutes of incubation. In contrast, both, no PS modification (#2, Figure 5), and single PS 3' of rC (#6, Figure 5 & 6) showed substantially higher resilience. Editing was also severely reduced for SNAP-ADAR1 and SNAP-ADAR2. It remains elusive as to how this drop in stability occurs, as one would expect that addition of more chemical modifications would increase stability rather than decreasing it. In contrast, when the gap was fully deoxy modified, no reduction of stability was detected when PS modifications were introduced as compared to the deoxy mismatch gap without PS. On the other hand, editing was again severely hampered (#12 versus #15, Figure 6A & B). Substituting the flanking deoxy nucleotides with 2'F only marginally increased editing but exhibited substantially lower protection in serum (#15, Figure 6 A & B).

Most intriguingly, the stabilizing effect of deoxynucleotides was transferable to other contexts outside of 5'-UAG. While the negative effect on editing was sometimes more severe for all tested SNAP-ADAR variants, it always severely boosted stability in 9 % FBS. For 5'-UAU and UAC guideRNAs with an unmodified mismatch gap were degraded instantaneously or within minutes (#16 & 20, Figure 7B & C). In contrast, deoxy modifications increased stability of for both to the maximum time-point of 24 h (#18, 19 & 22-25, Figure 7B & C). However, while editing was only slightly negatively impacted for 5'-UAC, it was more severely impacted for 5'-UAU, especially for editing with SNAP-ADAR1Q (Figure 7A). In contrast, editing of 5'-AAG was almost unchanged when introducing deoxy modifications while guideRNA stability was increased (#26 & 28, Figure 7). In addition, 2'OMe and PS modifications were also well accepted (#27 & 30, Figure 7). 2'F, on the other hand, caused a substantial drop in editing with the wildtype deaminase domains (#29, Figure 7). Intriguingly, 5'-CAC targeting guideRNAs exhibited maximum stability up to the maximum recorded timepoint of 24 h, which remained at that level when the mismatch C was flanked by 2'OMe modified nucleotides (#31 & 32, Figure 7B & C). In contrast, editing decreased strongly for all SNAP-ADAR variants (#31 & 32, Figure 7A).



C)

guideRNA	Context 5'-	Sequence
16	UAU	mC*mU*mA mGmGmC mAmAmC mA **rA** **rC** **rA** mUmCmC mAmC*mU* mU*mU*mA
17	UAU	mC*mU*mA mGmGmC mAmAmC mA **mA** **rC*mA** mUmCmC mAmC*mU* mU*mU*mA
18	UAU	mC*mU*mA mGmGmC mAmAmC mA **dA** **dC** **rA** mUmCmC mAmC*mU* mU*mU*mA
19	UAU	mC*mU*mA mGmGmC mAmAmC mA **dA** **dC** **mA** mUmCmC mAmC*mU* mU*mU*mA
20	UAC	mC*mC*mG mAmGmC mGmCmC mA **rG** **rC** **rA** mGmAmG mGmC*mA* mG*mG*mG
21	UAC	mC*mC*mG mAmGmC mGmCmC mA **mG** **rC*mA** mGmAmG mGmC*mA* mG*mG*mG
22	UAC	mC*mC*mG mAmGmC mGmCmC mA **dG** **dC** **mA** mGmAmG mGmC*mA* mG*mG*mG
23	UAC	mC*mC*mG mAmGmC mGmCmC mA **dG** **dC** **rA** mGmAmG mGmC*mA* mG*mG*mG
24	UAC	mC*mC*mG mAmGmC mGmCmC mA **dG** **dC*rA** mGmAmG mGmC*mA* mG*mG*mG
25	UAC	mC*mC*mG mAmGmC mGmCmC mA **dG** **dC** **rA*** mGmAmG mGmC*mA* mG*mG*mG
26	AAG	mU*mG*mU mAmUmA mUmCmC mA **rC** **rC** **rU** mUmAmC mCmA*mG* mA*mG*mU
27	AAG	mU*mG*mU mAmUmA mUmCmC mA **mC** **rC*mU** mUmAmC mCmA*mG* mA*mG*mU
28	AAG	mU*mG*mU mAmUmA mUmCmC mA **dC** **rC*dU** mUmAmC mCmA*mG* mA*mG*mU
29	AAG	mU*mG*mU mAmUmA mUmCmC mA **fC** **rC*fU** mUmAmC mCmA*mG* mA*mG*mU
30	AAG	mU*mG*mU mAmUmA mUmCmC mA **mC*** **rC*mU*** mUmAmC mCmA*mG* mA*mG*mU
31	CAC	mA*mA*mC mGmCmC mAmGmG mG **rG** **rC** **rG** mCmUmA mAmG*mC* mA*mG*mU
32	CAC	mA*mA*mC mGmCmC mAmGmG mG **mG** **rC*mG** mCmUmA mAmG*mC* mA*mG*mU

rN = RNA
dN = DNA
mN = 2'OMe
fN = 2'F
lN = LNA
N*N = PS

Figure 7: ***Stabilizing modifications are transferable to other contexts.*** *HEK 293 Flp-In™ T-REx™ cells stably expressing SNAP-ADARs variants were transfected with 5 pmol of snap-guideRNAs, respectively targeting various codon contexts within GAPDH transcript. Editing was assessed 24 h post transfection. Data shown as mean of ≥1 independent experiments, as indicated by points.* ***B)*** *guideRNAs were incubated in PBS containing FBS (9% final concentration) at 37 °C. At indicated timepoints 15 pmol of guideRNA was sampled. Samples were separated on 5M urea 20 % polyacrylamid gel and stained with SYBRGold.* ***C)*** *guideRNA sequence and mofication patterns. Nucleotides in bold represent the mismacth gap of various codon contexts within GAPDH transcript.*

With this data, we clearly demonstrate that deoxy nucleotides not only strongly confer resistance in serum but are also highly tolerated by all tested ADAR deaminase domains. What is more, we can show that this modification is readily transferrable to different contexts, indicating that it can be generalized. While it indeed effectively conveyed stability in 9 % FBS, its impact on editing varied in part upon the target codon itself. In future *in-vivo* applications, this might however be compensated by the strong increase in stability. With this in mind, extending incubation time of guideRNAs in serum, might be worth investigating, considering that translocation to the site of action in animal models might not be as easily achieved as in cell culture. Consequently, exposure to nucleases in serum is possibly substantially longer than 24 h.

In summary, the tested modification patterns lay the foundation for transfer to other, potentially disease-relevant targets. What is more, by providing necessary stability, other modifications that can mediate carrier-free delivery to target organs can be tested to effectively transfer targeted A-to-I editing from cell culture to animal models.

2.2.2. SNAP-ADARs efficiently edit murine methyl CpG binding protein 2

The following section contains results part of the Bachelor's thesis of Clemens Lochmann, which he performed under my co-supervision. Cloning of construct for creation of cell lines stably and constitutively expressing disease relevant targets and subsequent creation was performed by me. Design of constructs and conception of experimental set up was performed by me together with Thorsten Stafforst. All editing experiments performed in this section were executed by Clemens Lochmann for his Bachelor's thesis. All $(snap)_2$-guideRNAs except for #33 and #38 (both generated by me) were generated by Clemens Lochmann for his Bachelor's thesis.

Encouraged by our results when targeting various sites on GAPDH while strongly boosting guideRNA stability in serum, we sought to apply our learnings to a disease relevant target. To this end, we targeted R106Q (5'-CAA codon context) and W104X (5'-UAG codon context) in the murine methyl CpG binding protein 2 (mMeCP2)-enhanced green fluorescent protein (eGFP) transcript kindly provided by Gale Mandel's laboratory.[207] We stably integrated mMeCP2-eGFP bearing both mutations, respectively under constitutive expression in HEK 293 Flp-In™ T-REx™ cells expressing all above used SNAP-ADAR isotypes under doxycycline-induction, respectively.

Different from the guideRNA designs mentioned above, we employed a design established by Karthika Devi Kiran Kumar as part of her doctorate thesis. The following changes were applied.

First of all, guideRNA length was increased on either side of the mismatch C resulting in a 5'- (+3)9-C-12 design as opposed to the prior 5'- (+3)8-C-10 design. Secondly, at position 1, 3, 17, and 19 from the 5'-end of the antisense part, LNAs were employed. Other modifications remained as according to the design with unmodified mismatch gap described above. Lastly, a bivalent snap-substrate linker was attached to the 5'-ends from here on referred to as $(snap)_2$, which covalently binds two SNAP-ADAR moieties to the guideRNA and steers them to the target site. We further extended the incubation time of NH-guideRNAs in serum to one week.

Similar to the results obtained for 5'UAG editing on GAPDH, unmodified mismatch gap yielded highest editing for all tested SNAP-ADARs but showed miserable stability being instantly degraded in its entirety (#33, Figure 8A & B). Flanking a mismatch dC with 2'OMe modified nucleotides and one PS modification 3' of mismatch dC did not improve stability but retained editing capacity for all tested SNAP-ADARs (#34, Figure 8A & B). In contrast, fully deoxy modifying gap nucleotides provided sensational protection from degradation in serum showing strong bands even after seven days of incubation (#35 & 36, Figure 8B). Moreover, editing capacity of all tested SNAP-ADARs was only marginally reduced, which confirms the positive effects of deoxy modifications observed with guideRNAs targeting GAPDH (Figure 8A). What is more, all but one of the modified guideRNAs lack the 5' extension of non-binding nucleotides showing they are not relevant for SNAP-ADAR flexibility (Figure 8C). Lastly, we designed a guideRNA that contained a cholesterol modification at its 3'-end and two deoxy modifications in the mismatch gap (#37, Figure 8C). The cholesterol modification was reported to aid oligonucleotides to permeate the cellular membrane.[260] When transfecting nucleotides, permeation of the cellular membrane is usually mediated by Lipofectamine or similar reagents. With cholesterol those reagents might not be necessary severely facilitating future *in-vivo* application. Most intriguingly, while editing was slightly reduced as compared to its non-cholesterol bearing counterpart, stability in serum was extremely reduced from up to one week without cholesterol to 5 minutes with cholesterol (#35 versus 37, Figure 8A & B). It remains elusive as to how the cholesterol linker can impact stability to this degree. Considering that all data above suggested that the mismatch gap is the breaking point of

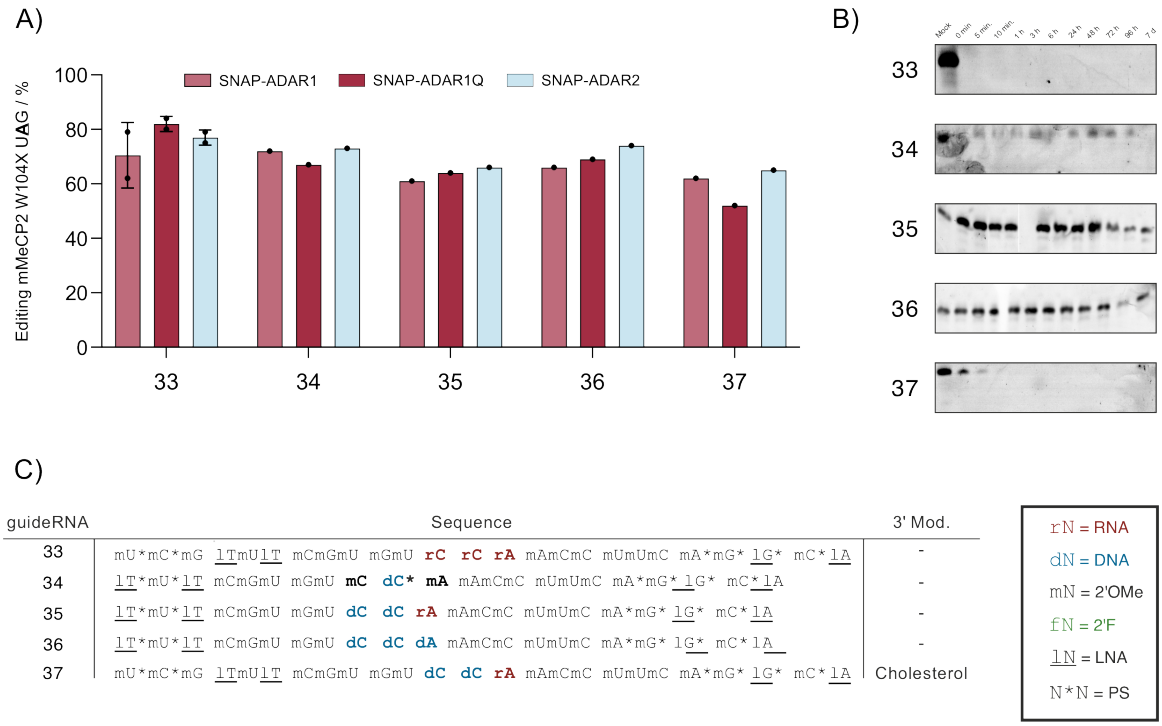


Figure 8: ***Deoxy modification maintains high editing levels in mMeCP2. A)*** *HEK 293 Flp-In™ T-REx™ cells stably expressing SNAP-ADARs variants and mMeCP2-eGFP containing W104X mutation were transfected with 1 pmol of (snap)$_2$-guideRNAs, respectively. Editing was assessed 24 h post transfection. Data shown as mean of ≥1 independent experiments, as indicated by points.* ***B)*** *NH-guideRNAs were incubated in PBS containing FBS (9% final concentration) at 37 °C. At indicated timepoints 15 pmol of guideRNA was sampled. Samples were separated on 5M urea 20 % polyacrylamid gel and stained with SYBRGold.* ***C)*** *guideRNA sequence and mofication patterns. Nucleotides in bold represent the mismacth gap of 5'UAG within mMeCP2W104X-eGFP transcript.*

the guideRNA, one would expect this guideRNA to be equally stable as without cholesterol.

It is possible that this phenomenon is simply an artefact caused by an interplay of the cholesterol linker with lipoproteins within the serum. Cholesterol-modified siRNAs were shown to interact with serum albumin and HDL/LDL.[260] It is possible that the cholesterol modified guideRNAs here associate with lipoproteins in the FBS and therefore do not pass through the polyacrylamide gel. Indeed, altered migration of cholesterol-modified oligonucleotides associated with LDL in polyacrylamide gels has been reported.[259] The fact that at 0 and 5 minutes of incubation a band is visible can be based upon slow kinetics of association with lipoproteins. This can be tested by proteolytic digest prior to loading onto the gel.

Next, we targeted the disease relevant R106Q mutation within mMeCP2-eGFP transcript in a 5'-CAA codon context (Figure 9). Again, deoxy modification severely boosted stability. Peculiarly, leaving a ribose guanosine 3' of the mismatch dC promoted higher protection than fully deoxy modifying the entire mismatch gap (#40 versus. #41, Figure

9B). As expected for 5'-CAN codons, editing was overall lower than for the 5'-UAG codon staying below 50 % for the hyperactive EQ variants and below 20 % for the wildtype variants with an unmodified guideRNA (#38, Figure 9A).

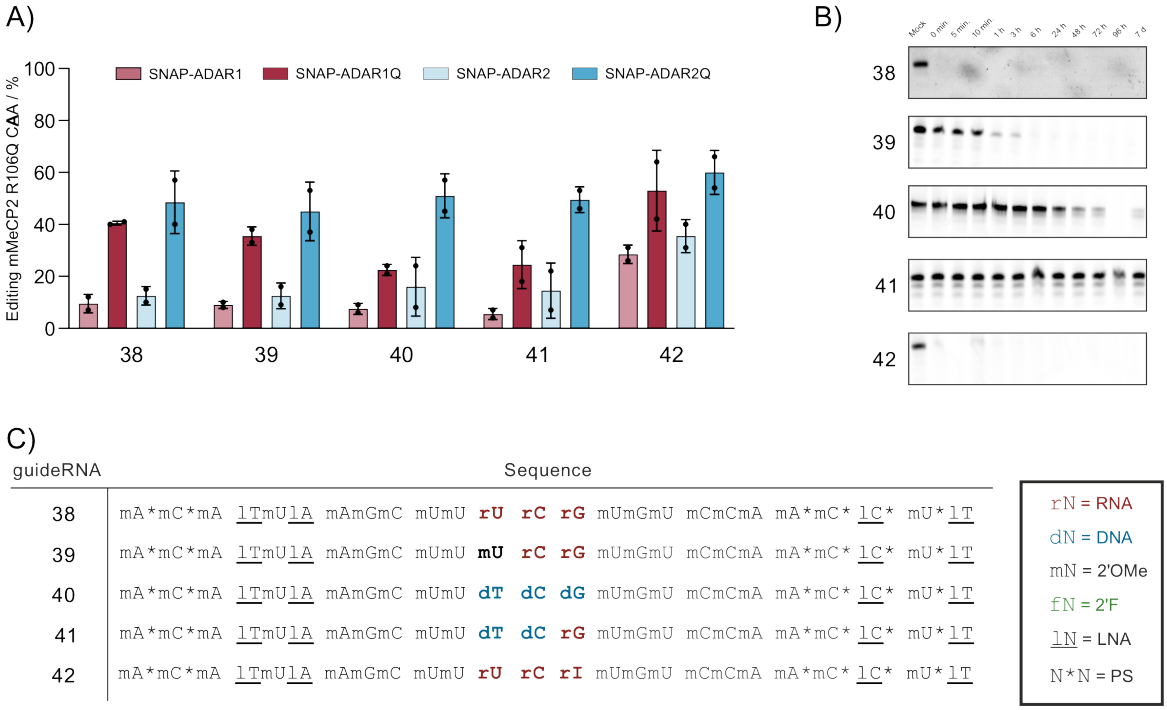


Figure 9: ***Chemical modifications boost stability and editing of 5'-CAA codon. A)*** *HEK 293 Flp-In™ T-REx™ cells stably expressing SNAP-ADARs variants and mMeCP2-eGFP containing R106Q mutation were transfected with 1 pmol of (snap)$_2$-guideRNAs, respectively. Editing was assessed 24 h post transfection. Data shown as mean of 2 independent experiments, as indicated by points.* ***B)*** *Cy5-guideRNAs were incubated in PBS containing FBS (9% final concentration) at 37 °C. At indicated timepoints 15 pmol of guideRNA was sampled. Samples were separated on 5M urea 20 % polyacrylamid gel and stained with SYBRGold.* ***C)*** *guideRNA sequence and mofication patterns. Nucleotides in bold represent the mismacth gap of 5'UAG within mMeCP2R106Q-eGFP transcript.*

Deoxy modifications further reduced editing most prominently for the SNAP-ADAR1Q variant. Editing levels of the other editases were either slightly lowered or even elevated for SNAP-ADAR2Q (#40 & 41, Figure 9A). As described in section 1.1.5, 5'-CAN codons are non-preferred by ADARs due to protrusion of guanine's amino group into the minor grove of the dsRNA and subsequent clash with one of the glycines flanking the strand invading glutamine/glutamate (Figure 2C).[142] Matthews and colleagues verified this by pairing the 5' C with an inosine, as hypoxanthine (inosine's base) lacks that amino group and showing elevated levels of editing (Figure 2C).[142] In accordance with that, we replaced the rG with an rI and report substantial increase for all tested SNAP-ADAR variants (#42, Figure 9A).

Taken together these results substantiate our stabilizing modification pattern to be readily transferable to other and even disease-relevant targets. What is more, for R106Q we demonstrated the true strength of the chemical modifications. By implementation of a non-

canonical nucleotide, we strongly enhanced editing of a non-preferred 5'-CAA context. This is a drawback of tools dependent on encoded guideRNAs. More importantly, this modification can also be utilized in oligonucleotides for targeted A-to-I editing with endogenous ADARs within the RESTORE system. Merkle et al. reported editing of the E342K mutation (5'-CAA context) within the serine protease inhibitor A1 causing α1-antitrypsin deficiency.[201] Here, Inosine opposite of the 5'-C could also strongly boost editing and thus possesses great therapeutic potential. Finally, with cholesterol we implemented a modification reported to mediate reagent-free uptake of siRNAs even *in-vivo*[260], with minimal impact on editing. This modification in particular can pave the way for future applications in animal models even beyond SNAP-ADAR mediated targeted A-to-I editing.

2.2.3. guideRNA design for agent-free delivery

The following section contains results part of the Bachelor's thesis of Clemens Lochmann, which he performed under my co-supervision. All editing experiments, generation of all $(snap)_2$-guideRNAs, and stability assays he performed. Design of constructs and conception of experimental set up was performed by me together with Thorsten Stafforst.

Stabilizing of the guideRNA was most ideal when deoxy modifications were applied in the mismatch gap. What is more, this modification also provided the most advantageous trade-off of stability versus editing impairment. Consequently, we sought to further expand modification patterns. In particular, we sought to establish a modification pattern that would further provide stability and assist in a reagent-free application of the guideRNA. For this, we introduced PS modifications, as they are reported to allow for gymnotic uptake of oligonucleotides.[221] We first employed PS outside the mismatch gap and then throughout the entire guideRNA (Figure 10D).

While stability in medium was high for even only partly PS-modified guideRNAs, full PS modification provided maximum stability of up to seven days in 9 % serum (#45 & 49, Figure 10B). Most intriguingly, however, was the fact that the impact on editing was only marginal even for the fully PS modified guideRNA (Figure 10A). While it had already been reported that PS modifications provide protection from nuclease degradation, the almost absence of editing impairment was noteworthy, as Matthews and colleagues have shown that multiple PO linkages are interaction points of ADAR's deaminase domain (Figure 2C).[142] On the other hand, PS modified ASOs have also been demonstrated to better interact

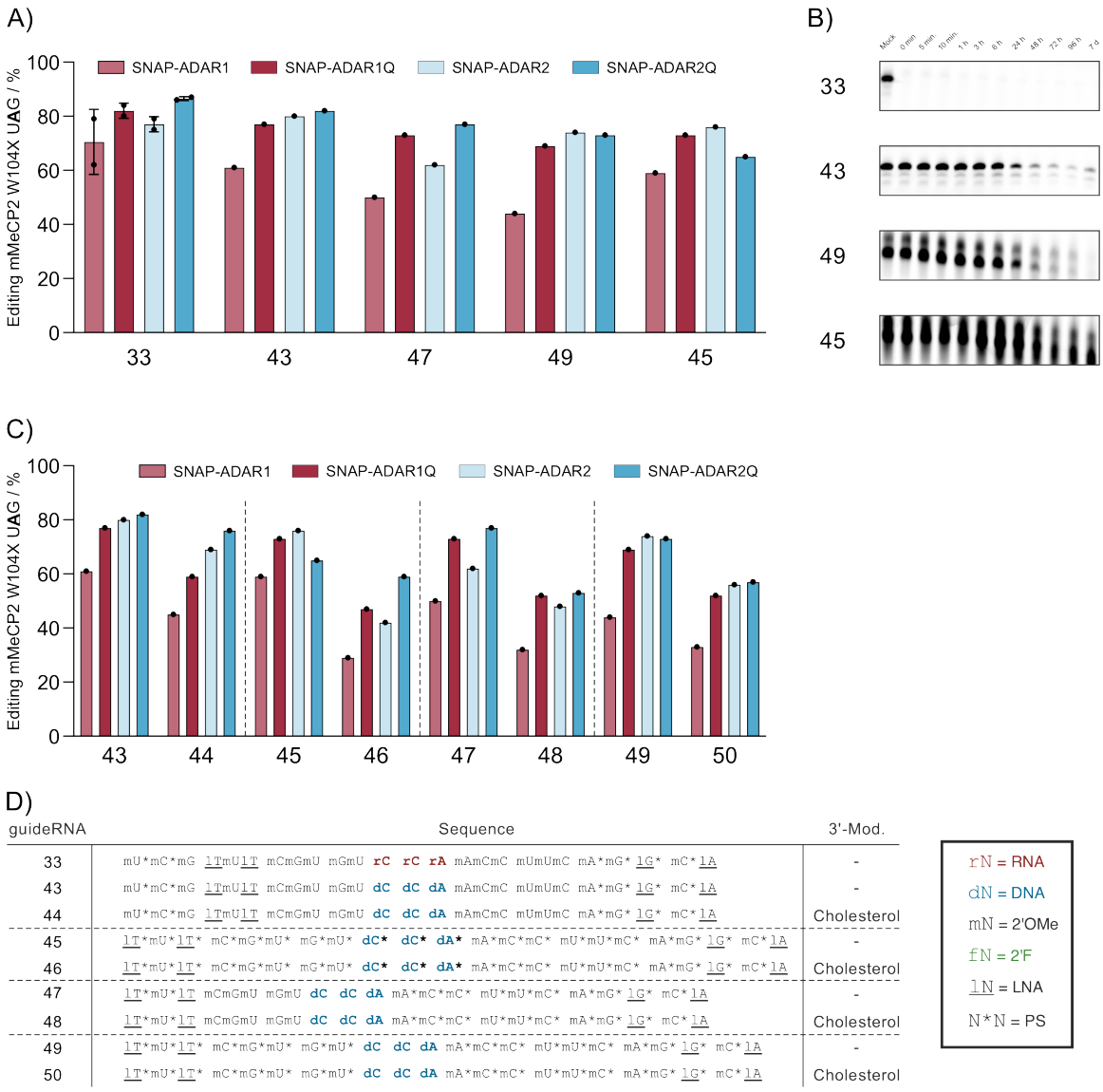


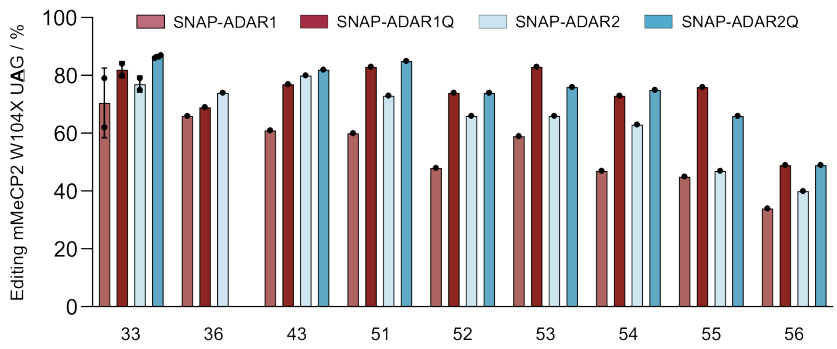
Figure 10: ***PS modification increase stability while preserving editing capacity. A)*** *HEK 293 Flp-In™ T-REx™ cells stably expressing SNAP-ADARs variants and mMeCP2-eGFP containing W104X mutation were transfected with 1 pmol of (snap)$_2$-guideRNAs, respectively. Editing was assessed 24 h post transfection. Data shown as mean of >1 independent experiments, as indicated by individual points.* ***B)*** *Cy5-guideRNAs were incubated in PBS containing FBS (9% final concentration) at 37 °C. At indicated timepoints 15 pmol of guideRNA was sampled. Samples were separated on 5M urea 20 % polyacrylamid gel and fluorescence was detected at 646 nm wavelength.* ***C)*** *FlpIn TReX cells stably expressing SNAP-ADARs variants and mMeCP2-eGFP containing W104X mutation were transfected with 1 pmol of (snap)$_2$-guideRNAs-Chol, respectively. Data shown as editing yield of 1 independent experiment, as indicated by individual points.* ***D)*** *guideRNA sequence and mofication patterns. Nucleotides in bold represent the mismacth gap of 5'UAG within mMeCP2R106Q-eGFP transcript.*

with proteins.[221, 223, 224] This might have a beneficial effect on dsRNA substrate recognition by SNAP-ADARs. Moreover, for RNase H-dependent ASO position-dependent stereopurity of PS modifications within the core region further increased efficiency than racemic PS modifications.[247, 248] Considering the interaction of ADAR's deaminase domain with the substrate dsRNA, this approach could even further boost editing yields for all applied SNAP-ADARs.[142] What is more, this modification pattern can be employed for

editing with endogenous ADARs within the RESTORE system, as PS modification patterns of the antisense part are similar to that employed for the SNAP-ADAR guideRNAs.[181, 182, 201]

Next, we further investigated cholesterol modification and designed guideRNAs with aforementioned PS modification pattern bearing cholesterol at their 3'-end (Figure 10D). For all tested PS modifications, cholesterol reduced editing levels with every SNAP-ADAR variant (Figure 10C). This can have various reasons. On the one hand, complex formation with the transfection reagent might be disturbed due to the lipophilic nature of cholesterol. On the other hand, the bulky modification itself might be obstructive for the deaminase domain. This could be verified in a cell-free *in-vitro* editing assay with purified target mRNA and SNAP-ADAR proteins.[180] Another possibility is a combined effect of PS modifications and cholesterol, changing proportions of productive and non-productive uptake of guideRNAs by association with cellular proteins or proteins in serum.

A)



B)

guideRNA	Design	Sequence
33	(+3) 9-C-12	mU*mC*mG lTmUlT mCmGmU mGmU **rC rC rA** mAmCmC mUmUmC mA*mG* lG* mC*lA
36	9-C-12	lT*mU*lT mCmGmU mGmU **dC dC dA** mAmCmC mUmUmC mA*mG* lG* mC*lA
43	(+3) 9-C-12	mU*mC*mG lTmUlT mCmGmU mGmU **dC dC dA** mAmCmC mUmUmC mA*mG* lG* mC*lA
51	8-C-11	lT*mU*lC mGmUmG mU **dC dC dA** mAmCmC mUmUmC* mA* lG*mG* lC
52	7-C-10	lT*mC*lG mUmGmU **dC dC dA** mAmCmC mUmU*mC* lA*mG*lG
53	6-C-9	lC*mG*lT mGmU **dC dC dA** mAmCmC mU*mU*lC* mA*lG
54	9-C-9	lT*mU*lT mCmGmU mGmU **dC dC dA** mAmCmC mU*mU*lC* mA*lG
55	8-C-8	lT*mU*lC mGmUmG mU **dC dC dA** mAmCmC* mU* lT*mC* lA
56	7-C-7	lT*mC*lG mUmGmU **dC dC dA** mAmC*mC* lT*mU*lC

rN = RNA
dN = DNA
mN = 2'OMe
fN = 2'F
lN = LNA
N*N = PS

Figure 11: ***Short guideRNAs retain editing capacity for all SNAP-ADARs. A)*** *HEK 293 Flp-In™ T-REx™ cells stably expressing SNAP-ADARs variants and mMeCP2-eGFP containing W104X mutation were transfected with 1 pmol of (snap)$_2$-guideRNAs, respectively. Editing was assessed 24 h post transfection. Data shown as mean of >1 independent experiments, as indicated by individual points.* ***B)*** *guideRNA sequence and mofication patterns. Nucleotides in bold represent the mismacth gap of 5'UAG within mMeCP2R106Q-eGFP transcript.*

Besides chemical modification, we considered size of the oligonucleotides to play a role, as it was reported for RNase H-dependent ASOs.[221] Therefore, we designed guideRNAs

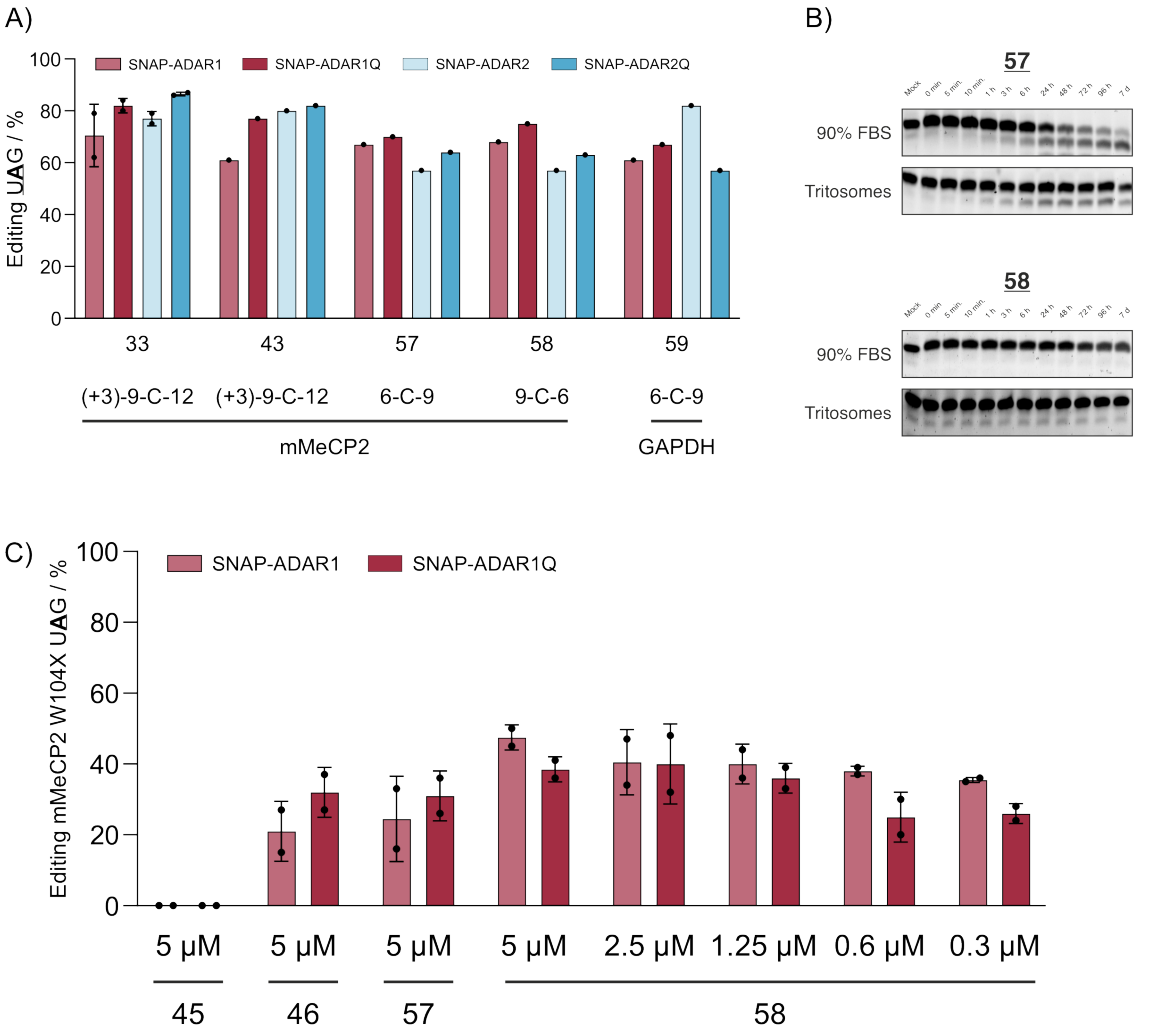
with deoxy modified mismatch gap and shortened them from both sides to maintain the asymmetric design and then tested a symmetric design as well (Figure 11C). Starting from the established 9-C-12 design we decrease the binding region down to 16 nucleotides with a 6-C-9 design. For the symmetric designs, we tested 9-C-9, 8-C-8, and 7-C-7. While shortening of the guideRNAs did indeed have an effect on editing, a more pronounced impact was observed when symmetries were altered. Here it seems clear that an asymmetric design is preferred by all tested SNAP-ADARs with 6-C-9 being the smallest one we deemed having the most acceptable reduction of editing (Figure 11A).

Finally, we combined all learnings to create an ideal guideRNA for carrier-free uptake for targeted A-to-I editing. We created a short 6-C-9 guideRNA design with a deoxy modified mismatch gap and with a fully PS modified backbone targeting mMeCP2W104X-eGFP and GAPDH, respectively. Besides 6-C-9, we included a reversed 9-C-6 design for mMeCP2W104X-GFP, which bore the $(snap)_2$-substrate at its 3'-end (Figure 12D). To further challenge stability of our designs, we decided to incubate guideRNAs in 90% FBS, as well as extracts from rat lysosomes referred to as Tritosomes.

As expected, editing yields remained at a very high level for all tested designs. Interestingly, reversing the guideRNA design from 6-C-9 to 9-C-6 did not negatively impact editing at all (Figure 12A).

In stark contrast, stability was severely impacted by this. While 6-C-9 was stable up to one week in Tritosomes, it showed drastic degradation after 24 h of incubation in 90 % FBS. On the other hand, 9-C-6 design seemed unaffected by either of the two mediums (Figure 12B). A simple explanation is not easily provided, as the modification pattern of both guideRNAs is the same and sequence differences are quite small. However, when closely examining the gel, it can be observed for guideRNA #57 (i.e. 6-C-9 design) that upon fainting of the intact band another band slightly below it becomes more prominent (Figure 12B, upper gel panel). We hypothesized that, while the guideRNA itself is very stable, maybe the C6-amino linker, to which the $(snap)_2$-substrate is conjugated is cleaved off by exonucleases present in the FBS. The absence of degradation product for the 9-C-6 guideRNA suggests 5' exonuclease activity in the serum (#58, Figure 12B, lower gel panel). In our application, this constitutes a major issue, as removal of the linker prevents the guideRNA to covalently bind to the SNAP-ADARs and steer them to the target site.

Next, we applied our guideRNAs for gymnotic uptake in a carrier-free delivery to cells expressing SNAP-ADAR1 and SNAP-ADAR1Q, respectively under doxycycline induction targeting stably integrated and constitutively expressed mMeCP2W104X-eGFP.



D)

guideRNA	5'-Mod.	Sequence	3'-Mod.
33	$(snap)_2$	mU*mC*mG lTmUlT mCmGmU mGmU **rC rC rA** mAmCmC mUmUmC mA*mG* lG* mC*lA	-
43	$(snap)_2$	mU*mC*mG lTmUlT mCmGmU mGmU **dC dC dA** mAmCmC mUmUmC mA*mG* lG* mC*lA	-
45	$(snap)_2$	lT*mU*lT* mC*mG*mU* mG*mU* **dC* dC* dA*** mA*mC*mC* mU*mU*mC* mA*mG* lG* mC*lA	-
46	$(snap)_2$	lT*mU*lT* mC*mG*mU* mG*mU* **dC* dC* dA*** mA*mC*mC* mU*mU*mC* mA*mG* lG* mC*lA	Cholesterol
57	$(snap)_2$	lC*mG*lT* mG*mU* **dC* dC* dA*** mA*mC*mC* mU*mU*lC* mA*lG	-
58	-	lT*mU*lT* mC*mG*mU* mG*mU* **dC* dC* dA*** mA*mC* lC*mU*lT	$(snap)_2$
59	$(snap)_2$	lT*mU*lT* mU*mU* **dC* dC* dA*** mG*mA*mC* mG*mG*lC* mA*lG	-

rN = RNA
dN = DNA
mN = 2'OMe
fN = 2'F
lN = LNA
N*N = PS

Figure 12: ***Gymnotic uptake dependent editing of mMeCP2W104X-eGFP. A)*** *HEK 293 Flp-In™ T-REx™ cells stably expressing SNAP-ADARs variants and mMeCP2-eGFP containing W104X mutation were transfected with 1 pmol of $(snap)_2$-guideRNAs, respectively. Editing was assessed 24 h post transfection. Data shown as mean of ≥1 independent experiments, as indicated by points.* ***B)*** *NH-guideRNAs were incubated in FBS (90% final concentration) at 37 °C. At indicated timepoints 15 pmol of guideRNA was sampled. Samples were separated on 5M urea 20 % polyacrylamid gel and stained with SYBRGold. C) FlpIn TReX cells stably expressing SNAP-ADAR variants and mMeCP2-eGFP containing W104X mutation were supplied with (snap)2-guideRNAs at indicated concentrations, respectively. Editing was assessed 72 h post supplementation. Data shown as mean of 2 independent experiments, as indicated by points.* ***D)*** *guideRNA sequence and mofication patterns. Nucleotides in bold represent the mismacth gap of 5'UAG within mMeCP2W104X-eGFP transcript.*

We tested guideRNAs fully PS modified and with a deoxy modified mismatch gap of 9-C-12 with and without cholesterol and 6-C-9 design (Figure 12D), respectively at a concentration of 5 µM. For 9-C-6 (#58) we tested concentrations from 5 µM down to 0.3 µM, as the stability data suggested that it shows higher resilience to harsh medium, hypothesizing that linker integrity is rather maintained than with the designs bearing the linker at their 5'-end.

While the short designs (#57 & 58) elicited editing at any applied concentration, the longer design relied on cholesterol for uptake into the cells (#45 versus #46, Figure 12C). Without cholesterol, 9-C-12 did not elicit any detectable editing (#45, Figure 12C). In contrast, when cholesterol was attached to its 3'-end, editing levels for both tested SNAP-ADAR1 variants rose to levels of 6-C-9 design (#46 versus 57, Figure 12C).

Interestingly, at applied concentration of 5 µM, 9-C-6 (#58) design elicited higher levels for both editases than 6-C-9 (#57), which is in accordance with our hypothesis that the 3'-end provides a more robust position for the linker than the 5'-end. Strikingly, all applied concentrations down to 0.3 µM yielded higher editing levels for SNAP-ADAR1 than with the other guideRNAs. In stark contrast, editing levels with SNAP-ADAR1Q were lower than with SNAP-ADAR1 for 6-C-9 design (#57, Figure 12C). This was unexpected and an explanation cannot be readily provided, as it was seen in the other experiments that in any case the hyperactive SNAP-ADAR1Q exceeded editing of the wildtype variant. An editing dependent toxicity due to increasing abundance of functioning mMeCP2 can be excluded, as transfection experiments yielding substantially higher editing did not exhibit this phenomenon.

Taken together these results demonstrate that we successfully expanded the modification patterns for SNAP-ADAR guideRNAs. In particular, PS-modifications further increased stability in harsh medium, while preserving editing capacity of all SNAP-ADARs. Intriguingly, it was possible to reduce guideRNA size from 22 nucleotides down to 16 nucleotides, while only marginally reducing editing yields. A combination of these two aspects resulted in a guideRNA design that we successfully applied for targeted A-to-I editing in reagent-free setting. However, editing levels were far lower than when guideRNAs were transfected. For this we transfected 1 pmol of guideRNA in 150 µl final volume (see Materials and Methods) resulting in a guideRNA concentration of about 6.7 nM. For the reagent-free application, we applied 45 (0.3 µM) to almost 750 (5 µM) times higher guideRNA concentrations. Only at 5 µM we obtained editing yields coming only remotely close to carrier-aided guideRNA delivery and only for one guideRNA design

(#58, Figure 12C). The other design elicited not even half the editing obtained by transfection. This data suggests that most of the guideRNAs we applied is either not taken up by the cells or is taken up and introduced to a non-productive pathway. In fact, this was not surprising, as it has been already reported that a considerably higher amount of oligonucleotides is required to elicit the same effect as a carrier-aided approach and most of applied naked oligonucleotides are probably taken up non-productively.[225, 249] Unfortunately, a quantitative ratio of total guideRNA in the cell to guideRNA in either productive or non-productive pathway cannot be easily provided, as the productive pathway would result in release of the guideRNA into the cytoplasm. Subsequently, signal of fluorescent tag might be diffused and thus not quantitatively captured.[225]

We further demonstrated that size of the oligonucleotide indeed impacts editing even when PS-modifications are applied. Here a longer guideRNA design could only elicit editing in a reagent-free uptake when cholesterol was attached to its 3'-end. This should be considered, as this was already reported for other oligonucleotide application to be pivotal even for delivery in animal models.[259-261] A combination of this with the shorter PS-modified guideRNAs could further boost editing. In addition, we observed a degradation pattern pointing to a destabilized linker. This can directly and severely impact guideRNA functionality. Lastly, here we applied guideRNAs once and analyzed editing 72 hours after. This setting can be optimized in multiple ways by duration of experiment or repetitive application of guideRNAs. This way degraded guideRNAs could be compensated, by which even the concentration of applied guideRNAs could be reduced.

In summary, we successfully determined a design for carrier-free delivery of guideRNAs in cell culture for targeted A-to-I editing. However, the low yield and high amount of employed guideRNA reveal plenty room for improvement for both the guideRNA design, and the application itself.

2.2.4. Linker stability in harsh medium

Within the laboratory rotation of Aline Maria Mack under my co-supervision, of which no data is shown here, we attempted to improve experimental set up for gymnotic uptake of our guideRNAs.

For this, we reapplied guideRNAs everyday over 72 hours keeping concentrations constant. However, for higher concentration of guideRNAs, this setting proved detrimental for cells, which is why editing yield could not be assessed. Following this, we increased the time

course of the experiment to seven days and reapplied guideRNAs once cells reached confluence. This setting was slightly better accepted by the cells but again editing yields could not be assessed or were at background level (data not shown).

In the light of stability data pointing to an unstable linker, these failed adjustments prompted us to more closely investigate the guideRNA stability in FBS and Tritosomes. In particular, we investigated the stability of linkers attached to either end of the guideRNAs. For this, we incubated NH-guideRNAs and $(snap)_2$-guideRNAs with 6-C-9 (#57) or 9-C-6 (#58) design, respectively in FBS and Tritosomes over seven days (Figure 13A).

In FBS, integrity of the 5'-NH/$(snap)_2$ guideRNA (#57) was compromised within 48 hours showing only a faint band after seven days of incubation. However, the bands slightly below the mock treated sample suggest that degradation of the guideRNA starts at the ends or at least in the immediate vicinity, as breaking points close to the center would yield substantially smaller bands (NH-57, Figure 13A, upper left panel). This becomes evident when this guideRNA bearing the $(snap)_2$-substrate is incubated in FBS (Figure 13A, upper left panel). The degradation pattern was similar to its NH-guideRNA counterpart. Within 48 hours most of the guideRNA was degraded showing only a very weak band after seven days of incubation. In contrast to the NH-guideRNA, the $(snap)_2$-guideRNA degradation product is most prominent at a size slightly below the intact NH-guideRNA (#57, Figure 13A, upper left panel). On the one hand, this in general proves that the side of degradation is indeed the linker bearing side. On the other hand, this result in particular suggests that the linkage between the C6-NH-linker at the 5'-end of the guideRNA and the $(snap)_2$-substrate is less likely affected than the linkage between C6-NH-linker and guideRNA itself. This might be caused by 5'-exonucleases present in the serum. This would also be in accordance with further smaller bands appearing over the course of the experiment suggesting further cleavage of 5'-terminal nucleotides (Figure 13A, upper left panel).

Interestingly, stability in Tritosomes was different. On the one hand, NH-guideRNA retained almost full integrity even after seven days with only a weak band for the degradation product. In stark contrast, the $(snap)_2$-guideRNA showed an even worse stability profile than in FBS with almost full degradation within six hours of incubation (#57, Figure 13A, lower left panel). Again, the degradation product was smaller than for the mock treated NH-guideRNA underpinning our hypothesis of $(snap)_2$-NH linkage being less affected. Most interestingly, the fact that this degradation is only seen upon conjugation with the $(snap)_2$-substrate suggests that the responsible enzymatic moiety in the Tritosomes does not recognize the guideRNA itself as a substrate. By extension, this suggests that this

is not caused by an exonuclease. The possibly responsible moiety might be the acetic phosphatase present in the Tritosomes.[253] The C6-NH-linker is attached to the first nucleotide of the guideRNA via a PO linkage. Consequently, upon conjugation of the $(snap)_2$-substrate to the C6-NH-linker the substrate might become large enough for it to be recognized by the acetic phosphatase and thus cleaved from the 5'-end of the guideRNA.

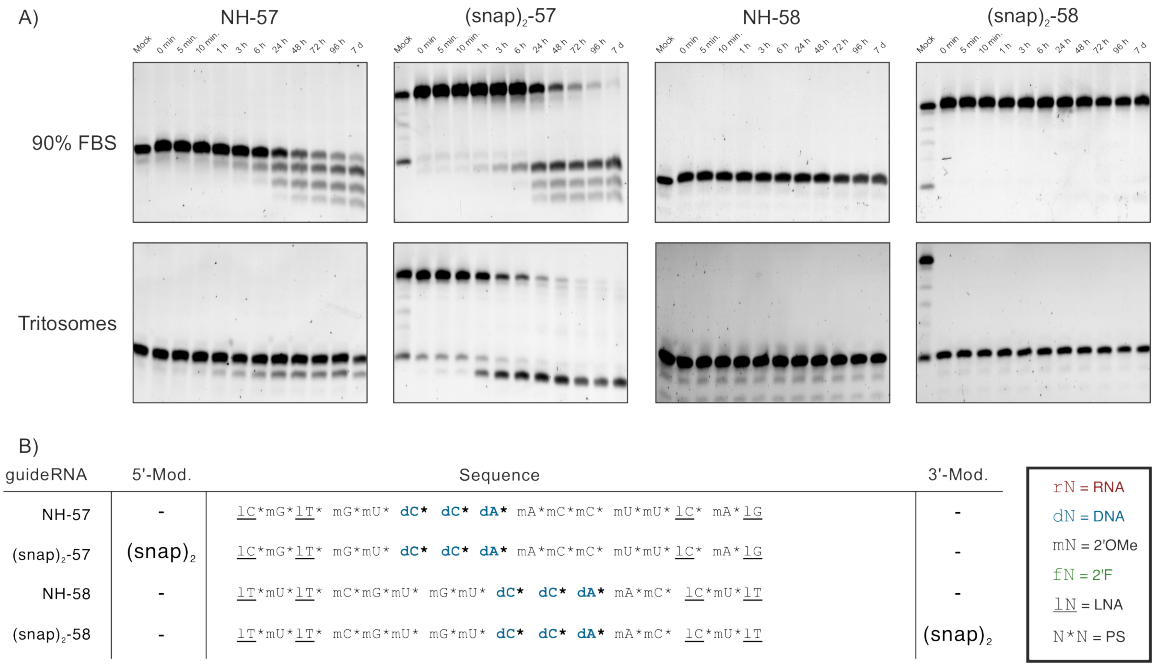


Figure 13: ***Linker integrity in harsh medium. A)*** *guideRNAs either bearing the $(snap)_2$-linker or only the NH-linker were incubated in FBS (90 % final concentration) or Tritosomes. At indicated timepoints 15 pmol of guideRNA was sampled. Samples were separated on 5M urea 20 % polyacrylamid gel and stained with SYBRGold.* ***B)*** *guideRNA sequence and mofication patterns. Nucleotides in bold represent the mismacth gap of 5'UAG within mMeCP2W104X-eGFP transcript.*

In the gymnotic uptake experiments 9-C-6 (#58) design with the $(snap)_2$-substrate at its 3'-end elicited higher editing levels than 6-C-9 (#57) with the $(snap)_2$-substrate at its 5'-end. We hypothesized that it is due to a higher linker stability in harsh medium. Indeed, in both, FBS and Tritosomes the NH-guideRNA retained full integrity over seven days of incubation, which was also observed for its $(snap)_2$-bearing counterpart incubated in FBS (#58, Figure 13A, right panel). In stark contrast, in Tritosomes, the $(snap)_2$-linker was instantaneously removed (#58, Figure 13A, lower right panel). Different from #57, degradation products here suggest that cleavage occurs between $(snap)_2$-substrate and the C7-NH-linker of the guideRNA. Following this, one is inclined to hypothesize that again no nuclease moiety is involved here, as 3'-exonucleases cleave 3'-PO from the sugar moiety, which in turn would mean that degradation products would run at a different height

than intact NH-guideRNA.[262] A definite answer for the breaking point can be obtained by mass spectrometric analysis of degradation products after incubation in Tritosomes.

Taken together these results neither confirm nor refute our hypothesis that linker stability is the basis upon which the 3'-end bearing guideRNA performed better in the gymnotic uptake experiment. Nevertheless, these data suggest that while all guideRNA with $(snap)_2$-substrate at its 3'-end would be degraded in a non-productive pathway, it is still substantially more stable before entering the cell even at 90 % FBS as compared to the guideRNA with the $(snap)_2$-substrate at its 5'-end. It is difficult to estimate to what degree this impacts our result, given the fact that cell culture medium used here contains only 10 % FBS. Irrespective of that, these data clearly show that linker stability is a weak point of our guideRNA design. Analogous to this, siRNAs require at PO at their guide strand's 5'-end for RISC loading, a feature that was recognized as week point of siRNA applications. This was addressed by employment of 5'-(E)-vinylphosphonate, a PO mimic that provided stability from phosphatases, while preserving efficient interaction with Argonaute 2 for RISC loading.[252] In fact, even a PS modification at that position had a positive effect, an approach that could easily be applied for $(snap)_2$-guideRNAs.[254]

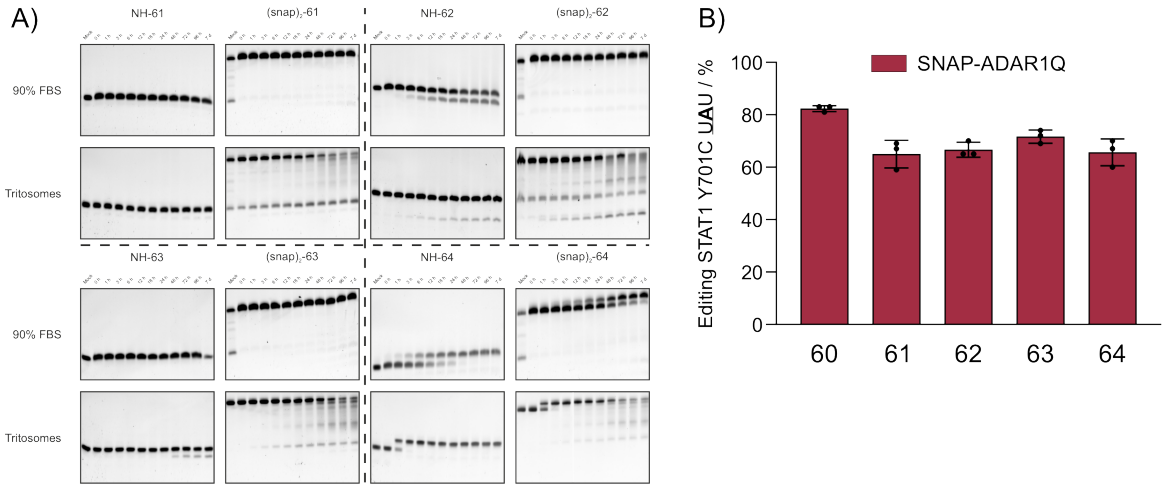
2.2.5. End modifications enhance linker stability

The following section contains results obtained by Clemens Lochmann during his laboratory rotation performed under my co-supervision. All editing experiments reported in this section were performed by Clemens Lochmann. All stability assessments reported in this section were performed by me. Design of constructs and conception of experimental set up was performed by me together with Thorsten Stafforst.

In the attempt of optimizing the gymnotic uptake setting for editing of mMeCP2 we encountered some difficulties. Firstly, application and reapplication of guideRNAs severely impacted cell viability (data now shown). Secondly, deeper assessment of guideRNA stability uncovered the stability of the $(snap)_2$-linker as a blind spot in our stability optimization efforts.

Consequently, we employed the following changes in our approach. Firstly, we designed guideRNAs with the C6-NH-/C7-NH-linker bound to a PS instead of a PO to the first/last nucleotide. In addition, we designed guideRNAs with an amino linker bound to a deoxy thymidine, which in turn is PS-linked (PS-dT) to the first/last nucleotide of the guideRNA. Secondly, we switched to HeLa cells constitutively expressing SNAP-ADAR1Q (created

by Karthika Devi Kiran Kumar), hoping that they show higher resilience to high amounts of guideRNAs. As the utilized HeLa cells do not express mMeCP2-eGFP, we targeted Y701 of endogenously expressed signal transducer and activator of transcription 1 (STAT1). However, we first assessed editing activity of those guideRNAs on HEK 293 Flp-In™ T-REx™ cells stably expressing SNAP-ADAR1Q under doxycycline induction. For the 5'-end bearing guideRNAs both modifications increased stability (#61 & 62, Figure 14A). In fact, a simple PS modification at the 5'-end helped retaining full linker integrity after seven days of incubation in FBS (#61, Figure 14A & C). In comparison, a guideRNA with a PO-linkage at the 5'-end lost almost all of its linker within 48 hours of incubation in FBS (#57, Figure 13A). The same was demonstrated for treatment with Tritosomes. Linker stability increased from 3 hours (#57, Figure 13A) to 24 hours with a simple PS-modification or 48 hours with a PS-dT modification at the 5'-end (#61 & 62, Figure 14A).



C)

guideRNA	5'-Mod.	Sequence	3'-Mod.
60	$(snap)_2$	mA*mG*mU lGmUlC mUmUmG mAmU **rA rC rA** mUmCmC mAmGmU mU*mC*lC* mU*lT	-
61	$(snap)_2$	*lT*mU*lG* mA*mU* **dA* dC* dA*** mU*mC*mC* mA*mG*lT* mU*lC	-
62	$(snap)_2$	**dT** *lT*mU*lG* mA*mU* **dA* dC* dA*** mU*mC*mC* mA*mG*lT* mU*lC	-
63	-	lG*mU*lC* mU*mU*mG* mA*mU* **dA* dC* dA*** mU*mC*lC* mA*lG*	$(snap)_2$
64	-	lG*mU*lC* mU*mU*mG* mA*mU* **dA* dC* dA*** mU*mC*lC* mA*lG* **dT**	$(snap)_2$

rN = RNA
dN = DNA
mN = 2'OMe
fN = 2'F
lN = LNA
N*N = PS

Figure 14: ***PS stabilizes linkers. A)*** *guideRNAs either bearing the $(snap)_2$-linker or only the NH-linker were incubated in FBS (90 % final concentration) or Tritosomes. At indicated timepoints 15 pmol of guideRNA was sampled. Samples were separated on 5M urea 20 % polyacrylamid gel and stained with SYBRGold.* ***B)*** *HEK 293 Flp-In™ T-REx™ cells stably expressing SNAP-ADAR variants were transfected with 1 pmol of $(snap)_2$-guideRNAs targeting Y701C within the endogenous STAT1 transcript, respectively. Editing was assessed 24 h post transfection. Data shown as mean of 3 independent experiments, as indicated by points.* ***C)*** *guideRNA sequence and mofication patterns. Nucleotides in bold represent the mismacth gap of 5'UAU within the endogenous STAT1 transcript.*

In similar fashion, linker stability of 3'-end bearing guideRNA was severely enhanced by both modifications (#63 & 64, Figure 14C). More importantly, stability in Tritosomes was

severely boosted (Figure 14A). While unmodified resulted in an instantaneous degradation of the linker (#58, Figure 13A), both modifications managed to retain linker integrity of more than 24 hours of incubation (#63 & 64, Figure 14A). Intriguingly, the PS-dT modified linker exhibited an unexpected degradation pattern (#64, Figure 14A, lower right panel). As degradation products are smaller than their native molecule, they run through the gel faster and thus appear as smaller bands in the gel. Here, degradation products run slower through the gel and appear as larger, albeit slightly (#64, Figure 14A, lower right panel). We hypothesized this to be due to the loss of PO at the 3'-end of the PS-dT. Consequently, upon cleavage caused by presumably acetic phosphatase electronegativity of oligo is reduced while retaining its size. As a result, the oligo runs slightly slower through the gel and thus appears slightly lager (#64, Figure 14A, lower right panel).

Nevertheless, linker stability is increased. Editing of Y701 in STAT1 with SNAP-ADAR1Q expressing HEK 293 Flp-In™ T-REx™ cells showed the usual slight reduction of editing, probably associated with the deoxy modified mismatch gap. Still editing levels were very high suggesting that these extra modifications are well accepted by SNAP-ADARs (Figure 14B).

Encouraged by these data, we sought to apply the guideRNAs for editing in HeLa cells stably and constitutively expressing SNAP-ADAR1Q (Figure 15). These cells were created by Karthika Devi Kiran Kumar. In order to assess the effect of linker stability on editing experiments, we tested editing kinetics of all end-stabilized guideRNAs in comparison with a guideRNA of ancestral (+3)9-C-12 design with unmodified mismatch gap (Figure 15B). For this, we transfected cells once with guideRNAs and measured editing yields every day for three days. Editing slightly dropped for the unmodified guideRNA but was still very high slightly below 80 % (#60, Figure 15A). This is something that was already described for HEK 293 Flp-In™ T-REx™ cells stably expressing SNAP-ADAR1Q. There, editing almost peaked three hours post-transfection and stayed relatively stable and slowly decreased after three days.[182] However, the end-modified guideRNAs exhibited a slightly steeper drop especially at 72 hours post transfection than for the unmodified guidRNA (Figure 15A). At first glance, a possible dilution effect caused by cell division seems unlikely as it would affect all guideRNA equally and Vogel et al. already reported that at similar editing levels in HEK 293 Flp-In™ T-REx™ cells this dilution effect started at day four post-transfection.[182] However, Vogel et al. also reported that editing with SNAP-ADAR1Q in HEK 293 Flp-In™ T-REx™ cells is marginally different if 5 pmol or 10 pmol of guideRNA are transfected and stays at slightly above 80% 24 hours post-transfection.[182]

This could suggest that, given mRNA turnover, editing is saturated at approximately 80 % and the guideRNAs are highly active so much so that even diluting them by cell division would not impact editing that severely. A way of testing that would be to transfect cells with lower amounts of guideRNA and test if the drop is steeper. Following this, a reduced editing capacity of the modified guideRNAs would then also lead to a steeper decline in editing upon dilution, as editing saturation is not reached. By extension this would suggest that the impact of the modifications on editing was more severe than can be seen here. On the other hand, given the fact that our guideRNA design remotely resembles RNase H gapmers with PS modifications throughout the oligonucleotide and a DNA gap, transcript stability should be taken into consideration and tested.[222-224] Indeed, while RNase H gapmers usually contain an 8-10 nucleotide large DNA gap, it was shown that even a gap of three DNA nucleotides can elicit RNase H-dependent degradation *in-vitro*.[230, 263] However, our guideRNAs here contain a mismatch in order to elicit targeted adenosine deamination and RNas H is reported to be sensitive to mismatches close to the cleavage site.[223, 264] In addition, our guideRNAs covalently bind two SNAP-ADAR moieties, which can also bind to the target mRNA and thus compete with RNase H. These aspects combined make it unlikely that the mRNA is degraded by RNase H.

Next, we applied these end-modified guideRNAs for gymnotic uptake in HeLa cells (Figure 15C). For this, we tested different concentrations of guideRNA by either applying once and incubating for three days or exchanging medium with fresh guideRNA-containing medium every day for three days. Editing yield increased in a dose-dependent manner with 1 µM showing highest editing for all set ups (Figure 15C). Surprisingly, overall daily reapplication did not increase editing and in some cases even reduced editing. An explanation is not readily provided, as we do not know what exact pathway the guideRNAs take to enter the cell. However, one aspect should be taken into consideration for this phenomenon. Prior to uptake, oligonucleotides adsorb to the cell membrane, which itself can be saturated and competitively inhibited with another oligonucleotide.[265, 266] To test if these HeLa cells are at all susceptible to repeated application of oligonucleotides, one could sequentially apply guideRNAs targeting different transcripts and assess respective editing levels.

As we could not increase editing by reapplication, we sought to keep it constant. For this, we applied fresh guideRNAs only when cells reached confluence and thus required passaging (Figure 15D, indicated timepoints). Here again editing remained at low levels but could be kept stable and even increased after multiple passages with the 5'-end bearing

guideRNA (#61, Figure 15D). For the 3'-end bearing guideRNA #63 editing peaked two days after the first passage and reapplication of the guideRNA and dropped at day seven again to editing levels obtained after initial guideRNA supplementation (Figure 15D).

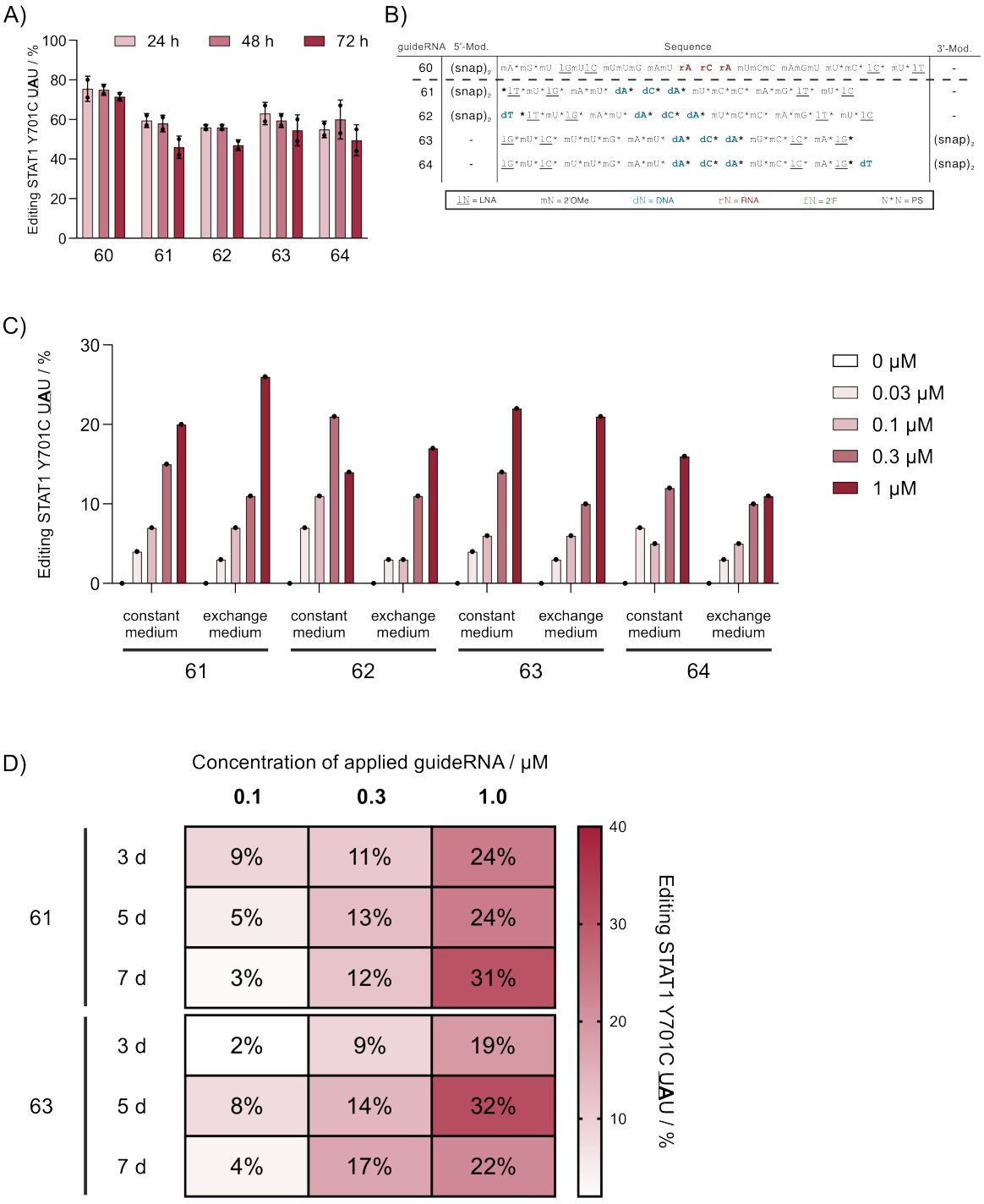


Figure 15: ***Carrier-free delivery of guideRNAs for STAT1 editing in HeLa cells. A)*** *HeLa cells stably expressing SNAP-ADAR1Q were transfected with 2 pmol of $(snap)_2$-guideRNAs targeting Y701C within the endogenous STAT1 transcript, respectively. Editing was assessed at indicated time-points post transfection. Data shown as mean of 2 independent experiments, as indicated by points.* ***B)*** *guideRNA sequence and mofication patterns. Nucleotides in bold represent the mismacth gap of 5'UAU within the endogenous STAT1 transcript.* ***C)*** *HeLa cells stably expressing SNAP-ADAR1Q were supplemented with $(snap)_2$-guideRNAs targeting Y701C within the endogenous STAT1 transcript, respectively at indicated final concentrations. Medium was either kept constant or exchanged with fresh medium containing $(snap)_2$-*

guideRNAs at indicated final concentrations. Editing was assessed 72 hours post initial guideRNA supplementation. Data shown as single independent experiment, as indicated by point. ***D)*** *HeLa cells stably expressing SNAP-ADAR1Q were supplemented with $(snap)_2$-guideRNAs targeting Y701C within the endogenous STAT1 transcript, respectively at indicated final concentrations. Medium was kept constant until the indicated time-points. At those time-points, cells reached confluence and were split. One part of the cells was utilized for editing assessment and the rest was given back to the well and were supplied with fresh $(snap)_2$-guideRNAs at indicated concentrations, 24 hours after splitting. Data shown as single independent experiment.*

A weak point of our application we addressed by modifying the $(snap)_2$-bearing ends of the guideRNAs with a PS. This did not just substantially increase stability in harsh medium but was well accepted by the SNAP-ADARs. At this point, it might be necessary to consider other modifications utilized by other oligonucleotide-based applications to protect from phosphatase activity. For siRNAs, a PO at the 5'-end of the guide strand is necessary for interaction of Argonaut 2 protein and thus for siRNA activity. Here, 5'-(E)-vinylphosphonate was reported to successfully mimic the PO linkage for loading onto RISC while severely boosting stability against phosphatases. With this modification singularly applied siRNAs retained activity *in-vivo* for weeks.[252]

However, while this modification might further increase stability of the linker and thus potentially increase duration of editing levels, a potentially major aspect might have been over-looked. In Figure 15A we can see that editing yield decreases even at high levels and even with very stable guideRNAs. As potential dilution effects or transcript stability should be taken into consideration, this reduction might also be based upon the nature of the utilized editases. More accurately, SNAP-tag is an enzyme derived from human AGT (see section 1.2.1), a suicide protein that recognizes alkylated guanines, transfers the alkyl to itself triggering its own proteolytic degradation.[169-172] It is possible that residual activity of the SNAP-tag remains, by which the SNAP-ADAR proteins are degraded upon binding to the guideRNA.

Like this, reserves of guideRNAs competent for covalent binding to SNAP-ADARs are depleted without actually degrading the guideRNAs themselves. A way to circumvent this would be to use HALO-tag instead. This mutant is derived from dehalogenase of *Rhodococcus rhodochrous*, an enzyme responsible for hydrolysis of haloalkanes within prokaryotic metabolism.[176] In contrast to SNAP-tag, substrate binding by HALO-tag does not elicit its proteolytic degradation and thus would not deplete functional guideRNA reserve. What is more, the linker itself is more hydrophobic than SNAP-tag substrate, which can have a beneficial effect on carrier-free delivery of guideRNAs.[169-172, 176]

Here, we can elicit stable, albeit rather low, editing over days when reapplying guideRNAs fully PS modified. As various application regimen exhibited similar outcome with a dose-

dependent editing yield, it might be interesting to take cholesterol or other lipophilic conjugates into consideration. As we demonstrated, cholesterol as a conjugate to a guideRNA without PS modification does indeed mediate gymnotic uptake for successful editing. Hence, a combination with shorter and PS-modified guideRNAs constitutes a promising option to further increase editing yield.

In conclusion, the synthetic nature of SNAP-ADAR guideRNAs allowed for extensive modification that not only mediate formidable long-term stability in very hostile environment but also allowed for carrier-free delivery and increased editing of difficult codons. More importantly, employment of functional molecules such as cholesterol or GalNAc would allow targeted delivery to specific tissues. What is more, all of these applied modifications can be employed within the RESTORE system, which utilized chemically modified oligonucleotides to elicit targeted A-to-I editing by recruitment of endogenously expressed ADARs.[201] While these modifications will very likely increase stability of RESTORE oligonucleotides and allow for their carrier-free delivery, the impact on editing is difficult to anticipate considering the involvement of the dsRBDs of the endogenous ADARs on target recognition. Lastly, acceptance of various sizes of guideRNAs, in addition to chemical modifications also close to the substrate linker make SNAP-tag itself an excellent platform for targeted A-to-I editing. This paves the way for expansion of the RNA editing field by employment of other effectors to elicit various effects. By combination with other self-labeling enzymes, this could create a platform with unmatched versatility and potential.

2.3. Targeted C-to-U editing with Apobec1-SNAP

2.3.1. murine Apo1-SNAP edits reporter effectively

In the beginning of this project, we sought to explore possibilities for expanding the available RNA editing platforms, which to that point were all limited to A-to-I editing, by engaging a natural cytidine deaminase for targeted C-to-U editing. Preliminary experiments, which are not part of this thesis, showed that when editing the endogenous ApoB mRNA target sequence within a dual reporter transcript (i.e., GFP-ApoB-mCherry kindly provided by Nina Papavasiliou laboratory), murine APOBEC-1 (mApo1) exhibited higher specificity and potency than its human counterpart. Therefore, we designated mApo1 to be our favorable deaminase and fused it to the amino terminus of the SNAP-tag, terming it mApo1-SNAP (mApo1S). Cloning, initial testing on GFP-ApoB-mCherry, and creation of HEK 293 Flp-In™ T-REx™ cells stably expressing mApo1S under doxycycline induction was performed by Madeleine Dias Mirandela (née Heep).

We initially tested a guideRNA bearing two SNAP-tag recruiting moieties (i.e., two benzylguanine), rationalizing this to be the best approach given endogenous ApoB mRNA editing is carried out in a complex containing an APOBEC-1 dimer (Manuscript 1, Figure 1A).[85, 86]

In order to deduce a rational guideRNA design, we targeted an over-expressed eGFP transcript with a $(snap)_2$-guideRNA that had been utilized before in monovalent (i.e. snap-guideRNA) fashion for editing of a 5'- UAG codon within the eGFP transcript using SNAP-ADARs.[183] As this kind of approach was unprecedented, we sought to explore if mApo1S could edit any surrounding cytidines. Interestingly, one cytidine within the eGFP transcript 7 nucleotides downstream of the binding site of the guideRNA within a 5'- ACG codon (T63) showed low levels of editing thereupon we designed further guideRNAs approaching the editing sites by one nucleotide at a time from the 5'-end and 3'-end, respectively (Manuscript 1, Figure 1B). While no guideRNA yielded any editing binding downstream of the target C, this experiment demonstrated that distance to it upstream was pivotal with 5 nucleotides constituting the ideal distance (Manuscript 1, Figure 1B). Hypothesizing that this distance preference might be due to ideal positioning with respect to the target C, we tested if imperfect distance could also be compensated by an increased antisense part (Manuscript 1, Figure 1C). By increasing the binding part by two nucleotides to 21 at the 3'-end of the guideRNA with a not ideal distance of 4 nucleotides to the target C, we could also achieve the same boost in editing (Manuscript 1, Figure 1C). Most

interestingly, combining putatively ideal distance and increased binding site, did not have an additional effect on editing yield. What is more, when targeting a different codon within the same transcript, that is glutamine (Q95) in a 5'- ACA context (Cave: the context does not necessarily reflect the coding base triplet for the targeted amino acid), we designed a guideRNA with a non-ideal distance and short binding site (19 nucleotides) but exchanged every other nucleotide of the guideRNA with LNAs. This only exhibited low levels of editing (Manuscript 1, Figure 1E).

While we successfully managed to create a SNAP-tagged cytidine deaminase for targeted RNA editing and managed to narrow down ideal guideRNA designs, transfer to another target within the same transcript did not have the desired effect. The endogenous ApoB target, as well as endogenous target sites of APOBEC-3A and -3G from secondary structures to recruit necessary proteins for efficient editing.[60, 118] This could explain discrepancies in those editing yields. On the other hand, it is known that APOBEC-1 translocates from a cytoplasmatic editing-inactive complex to the nucleus for editing.[38-41] Consequently, changing location of Apo1S might boost editing of unfavorable target sites. Finally, editing was analyzed 24 hours post-transfection. An increase of experiment duration may further boost editing.

2.3.2. mApo1S edits in the cytoplasm in a time-dependent manner

The following section discusses the results that are part of the bachelor's thesis by Aline Maria Mack carried out under my co-supervision. Cloning of constructs for creation of stable cells lines and subsequent creation were conducted by Aline Maria Mack. Design of constructs and conception of cloning and experimental plan was performed by me together with Thorsten Stafforst.

In order to increase editing yield with Apo1S, we considered assessing the degree to which its localization and enzymatic rate contribute to targeted C-to-U editing. To this end, in an extensive approach, Aline Maria Mack as part of her bachelor thesis created HEK 293 Flp-In™ T-REx™ cell lines stably expressing either NLS- or NES-tagged mApo1S constructs under doxycycline induction. In addition, we investigated if linkers between those functional peptides help by providing more flexibility (Manuscript 1, Supporting Figure S1, S2 & S4). While the latter did not have any effect, the tags demonstrated their functionality. The native Apo1S was mainly located within the nucleus but not exclusively, as signal was also detectable in the cytoplasm confirming reports showing its presence in

both compartments (Manuscript 1, Figure 1D, Supporting Figure S1, S2 & S4).[84-88] When the Simian Virus 40 (SV40) NLS was added to mApo1S, all protein was localized in the nucleus indicating that the putative NES/cellular retention signal (CES) of mApo1 was rather weak.[84-88] On the other hand, this could be due to the SNAP-tag being fused to mApo1's C-terminus, wherein the putative NES/CES lies.[84-88] Nevertheless, editing yields dropped strongly for every target site suggesting that the place of editing of mApo1S is the cytoplasm (Manuscript 1, Figure 1E, Supporting Figure S5). This contradicts APOBEC-1 activity on endogenous ApoB transcript. A proposed model of APOBEC-1 editing describes localization as a regulatory mechanism. APOBEC-1 is considered to be localized in a cytoplasmatic editing inactive higher order protein complex and localizes to the nucleus to an editing active smaller protein complex.[38-41] Again, the C-terminus is described to be involved in protein-protein interactions might be blocked by the SNAP-tag thus possibly explaining this outcome.[85, 86] However, considering that the main portion of native mApo1S is located in the nucleus (Manuscript 1, Figure 1D, Supporting Figure S1), this would imply that by adding an NES to the enzyme, editing should be increased. Surprisingly, editing was virtually unchanged as compared to the native enzyme, even though localization was entirely shifted to the cytoplasm ultimately demonstrating mApo1S behavior is completely uncoupled from endogenous APOBEC-1 (Manuscript 1, Figure 1D & E, Supporting Figure S2 & S3).

Another aspect we considered was enzyme activity and increased experiment time to 48 and 72 hours, respectively. Strikingly, this had the greatest impact on editing, substantially increasing editing for all edited sites (Manuscript 1, Supporting Figure S6). Since, the difference of 48 to 72 h was only marginal, we deemed 48 h to be sufficient for our purposes.

Taken together, these results indicate that targeted cytidine deamination with Apo1S is a cytoplasmatic process probably requiring only a fraction of the enzyme to take place and that is slower than adenosine deamination carried out by SNAP-ADARs.[182] Even though editing could not be enhanced when translocating mApo1S exclusively to the cytoplasm, we decided to carry on with that construct. The reason for this is reports that show that APOBEC-1 can also deaminate cytidines within a DNA molecule.[102, 103] To prevent possible adverse effects of DNA-editing we continued mRNA editing with NES-tagged mApo1S for editing of endogenous targets.

2.3.3. mApo1S exhibits poor editing capacity on endogenous transcripts

With mApo1S showing effective editing on eGFP when editing was carried out over 48 h, we sought to transfer this approach to editing of endogenous transcripts. For this, we designed gRNAs with the appropriate design and chemistry (i.e., 5 nucleotides distance, and 2'OMe and LNA) and targeted different sites on GAPDH, β-glucuronidase (GUSB), β-Actin (ActB), and peptidyl-prolyl cis-trans isomerase B (PPIB), respectively (Manuscript 1, Figure 1F).

For ActB, we targeted a tyrosine (Y166) to cause a silent mutation in a 5'- ACG context and a glutamine (Q49) causing a STOP codon in a 5'- UCA context. For GUSB, we also targeted two glutamines (Q277 and Q292), respectively, both in a 5'- UCA context. In GAPDH, we targeted a threonine (T52) in a 5'- ACC context to an isoleucine. It is worth mentioning that the extended site of the target contained multiple cytidines, by which we intended to test accuracy of mApo1S when supplied with multiple potential substrates (Manuscript 1, Supporting Figure S7). Given mApo1S can edit even with suboptimal positioning, at least two of those cytidines should be edited if our mApo1S lacked accuracy. This is particularly interesting when considering that our guideRNA only provides correct positioning by distance to the target base and does not mark it, as it is the case for ADAR applications (target adenine is in a mismatch with a cytidine).[180-182, 201, 209]

Unfortunately, while mApo1S elicited some degree of editing on GAPDH T52, none of the other targets exhibited any editing (Manuscript 1, Figure 1F). We hypothesized that low yield for the Q to Stop editings on GUSB and ActB to be due to non-sense mediated decay. In fact, the required distance to exon-exon junctions was given for these targets.[267] However, given the failed editing on the other codons, it is more likely that the editor exhibited insufficient potency on endogenous targets.

Taken together, we were able to establish a tool for targeted C-to-U editing using a SNAP-tagged endogenous deaminase. Unfortunately, while it exhibited extraordinary editing activity on an overexpressed reporter transcript, only one endogenous target could be edited at very low levels. Revisiting APOBEC-1's endogenous target, the secondary structure of ApoB mRNA is of central importance for efficient editing.[130] Something similar was reported for APOBEC-3A and APOBEC-3G.[60] This was exploited for the Cas13-based programmable cytidine deaminase CURE. Here, the guideRNA forces the target mRNA to form a loop, within which the target C is positioned for editing.[195] In fact, we tested this

kind of guideRNA design with mApo1S, but it did not elicit editing (data not shown). In conclusion, APOBEC-1 is an effective cytidine deaminase however its employment for targeted and programmable C-to-U editing is more complicated than anticipated. Even though we elicited higher editing on overexpressed reporter transcripts than other reported APOBEC-1 fusion proteins, the difficulty in controlling mApo1S renders it a suboptimal tool for effective targeted C-to-U editing.[194]

2.4. Orthogonal Recruitment of RNA Editors

2.4.1. Orthogonal recruitment of ADAR deaminase domains

The SNAP-tag constitutes a highly potent platform for targeted A-to-I editing, providing means to recruit both catalytically active ADAR deaminase domains and their respective hyperactive variants. Analysis of codon scope provided some insight into tolerance of different flanking nucleotides for the target A. Intriguingly, there were some complementarities of those two ADAR deaminase domains, that is, some codons were better edited by SNAP-ADAR1Q than SNAP-ADAR2Q and vice versa.[182] This generated a thought-provoking impulse to expand the effector fusion platform to provide a possibility to engage different editors simultaneously to better control codon preferences.

For this reason, Anna Sophia Imrich (née Stroppel) as part of her doctoral thesis extensively tested different self-labeling enzymes ultimately for orthogonal recruitment of different editases for performing simultaneous editing on different targets. This orthogonality was provided by conjugating the according linker to the guideRNA. As I did not contribute to results described in this section, I will briefly summarize key findings and results. Experiments to which I contributed are described in the next section.

HEK 293 Flp-In™ T-REx™ cell lines stably expressing SNAP-, CLIP-, or HALO-tagged ADAR deaminase domains under doxycycline induction were transfected with guideRNAs bearing respective linkers to edit different transcripts with the according enzymes. Orthogonality was challenged by transfection with non-substrate guideRNAs (Manuscript 2, Figure 1D, Supporting Figure S11, S12, S13A-D). Accordingly, SNAP-ADARs would edit only when supplied with snap-gRNAs, CLIP-ADARs with clip-gRNAs, and HALO-ADARs with halo-gRNAs. The initial test on an over-expressed eGFP transcript revealed that sufficient orthogonality could only be achieved when using SNAP- and HALO-ADARs (Manuscript 2, Figure 1D, Supporting Figure S11 & S12). With SNAP-tag being the ancestor of CLIP-tag, lack of selectivity of those enzymes was not surprising.[175] In addition, clip-gRNAs exhibited severe instability upon long-term storage (Manuscript 2, Supporting Figure S12).

With SNAP- and HALO-tag, it was possible to recruit either ADAR1Q or ADAR2Q deaminase domain with maximum selectivity. This way in cells stably expressing both, HALO-ADAR1Q, and SNAP-ADAR2Q this selectivity extended the codon scope we could address (Manuscript 2, Figure 2 & 3, Supporting Figure S13A-D). Vogel et al. had shown before that deaminase domains of both ADARs have slightly different codon

preferences.[182] With this system we were able to edit every preferred codon, respectively. In fact, designing a guideRNA with a bi-functional linker both deaminase domains could be recruited simultaneously with comparable results to editing with a mono-functional linker (Manuscript 2, Figure 3G & H). What is more, transcriptome-wide off-target analysis of this dual expressing cell line indeed showed an increase in off-targets as compared to cells stably expressing only one editase (Manuscript 2, Table 1, Figure 4, Supporting Figure S16). However, the number of off-target sites for the dual expressing cell line amounted to an approximate accumulated number of off-target of both single expressing cell lines (Manuscript 2, Table 1). In addition, as described before for the SNAP-ADARs, by far most of the off-target sites were caused by the over-expressed editases themselves unaffected by guideRNA fidelity.[182] What is more, for about 75 % of those sites editing yield was below 25 %. Of the guideRNA-dependent off-target sites, only 37 showed editing levels above 25 %, of which only five were missense (Manuscript 2, Supporting Table S6). All of those 37 sites showed sequence homology to either of the utilized guideRNA (Manuscript 2, Supporting Table S7, Supporting Figure S17-S19).

With the HALO-tag we successfully and effectively augmented the SNAP-ADAR platform for targeted A-to-I editing with both ADAR deaminase domains for extended codon scope. In fact, by implementation of chemical modifications, editing levels can be even more boosted. As described above, inserting inosine opposite of a C in a 5'-CAA context, editing with all SNAP-ADAR variants was strongly increased. Moreover, chemical modifications could also help reduce transcriptome-wide off-targets, mainly caused by over-expressed hyperactive ADAR-EQ variants. By incorporating Banner's base at the mismatch C position, on-target editing levels with the wildtype deaminase domain could be as high as with the hyperactive EQ variants while not sharing their off-target profile.[229] In summary, the establishment of this orthogonal platform paves the way for applications beyond targeted A-to-I editing.

2.4.2. Orthogonal A-to-I and C-to-U editing

To expand our repertoire of effector proteins Anna Sophia Imrich (née Stroppel), as part of her doctoral thesis, created a HEK 293 Flp-In™ T-REx™ cell line stably expressing both HALO-ADAR1Q and mApo1S under doxycycline induction (Manuscript 2, Figure 5A). We first targeted a 5'-UAG codon with HALO-ADAR1Q and a 5'-ACG codon with Apo1S in eGFP. We effectively edited both sites with either one guideRNA for each site bearing

the respective substrate-linker or with one guideRNA for both sites with a bi-functional linker (Manuscript 2, Figure 5C & D). The latter was possible, as we established a guideRNA design that positions Apo1S five nucleotides upstream of the target C (section 2.3.1). Further tests on endogenous transcript (ActB for HALO-ADAR1Q and GAPDH for mApo1S) demonstrated that it is effectively possible to edit different endogenous target sites with both editases simultaneously (Manuscript 2, Figure 5F).

My contribution consisted of performing editing experiments for transcriptome-wide off-target analysis of the dual expressing cell line for HALO-ADAR1Q and mApo1S and the analysis of the nature of induced mutations, as well as benchmarking experiments with the Cas13-based RESCUE system.

For the latter, Moritz Stoll created a HEK 293 Flp-In™ T-REx™ cell line stably expressing RESCUE under doxycycline induction. Design of constructs and conception of cloning and experimental plan were performed by together with Thorsten Stafforst. Editing experiments with those cell lines were performed by me.

We targeted both eGFP T63 (5'-ACG context) and GAPDH T52 (5'-ACC context). As by author's suggestions, we designed multiple guideRNA for RESCUE with different mismatch C positions to establish the most potent design (Manuscript 2, Supporting Figure S15).[191] This benchmark revealed that mApo1S outperformed RESCUE for the tested target sites. For the eGFP target RESCUE elicited low levels of about 10 % for every guidRNA design. In contrast, mApo1S editing of over 60 % with low levels of by-stander C-to-U editing on two sites (Manuscript 2, Supporting Figure S15A). For GAPDH almost every RESCUE guideRNA design elicited low levels of on-target editing but also elicited equal levels of both A-to-I and C-to-U by-stander editing on multiple sites. On the other hand, mApo1S exhibited low-level editing on one by-stander C and high-level editing (higher than on-target) of another by-stander C (Manuscript 2, Supporting Figure S15B). In fact, this by-stander we also observed when editing with mApo1S alone and we hypothesize that the reason for this is a predicted stem-loop positioning that by-stander C at the edge of the loop (Manuscript 1, Supporting Figure S7), similar to other APOBEC-1 endogenous target sites.[130]

Transcriptome-wide off-target of both A-to-I, and C-to-U editing of our dual expressing cell line revealed a higher amount of detected A-to-I off-target sites than for the dual expressing cell line expressing both HALO-ADAR1Q and SNAP-ADAR2Q (Manuscript 2, Table 1 & 2). This can be attributed to the fact that editing with the A-to-I and C-to-U editing cells was analyzed 48 hours after guideRNA transfection due to the lower enzyme

activity of mApo1S (section 2.3.2). However, again by far most of the sites exhibited editing levels lower than 25 % (Manuscript 2, Supporting Table S10 & 11). Of those above 25 % A-to-I editing only 383 sites lead to a change in coding sequence (Manuscript 2, Supporting Table S10). Again, only a small part of the off-target sites with over 25 % editing was guideRNA-dependent (129 sites in total, 49 sites with change in coding sequence).

However, most intriguingly, mApo1S only elicited very few C-to-U off-target editing sites with editing above 25 % with 129 sites in total, 109 of which located in 3'UTR and only three guideRNA-dependent (Manuscript 2, Supporting Table S11). For sites with an editing level above 10 %, we detected more sites (1009 sites). Again, only a fraction (44 site) was guideRNA-dependent and by far most of the sites were located in the 3'UTR (Manuscript 2, Supporting Table S11), as was expected from reports showing that most APOBEC-1 editing sites are located in 3'UTRs.[109, 129]

Taken together, these results show that orthogonality of HALO- and SNAP-tag can be exploited to elicit not just A-to-I editing of different codons with ADAR deaminase domains but also C-to-U editing with mApo1S. Despite mApo1S's transcriptome-wide off-targets low in both number and editing yield and effective editing of demonstrated target sites, even outperforming the RESCUE system, known caveats described above remain. On the one hand, low enzymatic activity of mApo1S required us to double incubation time as compared to editing with ADAR deaminase domains, by which off-target editing sites of HALO-ADAR1Q increased substantially. On the other hand, difficulty in transferring guideRNA design rules to other endogenous targets and a reported rather limited codon scope of 5'-ACN codons render mApo1S a suboptimal tool.[129] In contrast, while RESCUE did not elicit satisfactory editing for the tested targets it still holds potential. For one, RESCUE is a fusion protein of Cas13 and a mutated deaminase domain of ADAR2 and exhibits a very broad codon scope.[191] By substituting the Cas13 protein with a SNAP-tag a better editase for targeted C-to-U editing could be created.

In conclusion, a combination of HALO- and SNAP-tag can lay the foundation for a platform for employment of a multitude of different effectors that can modulate target RNA fate. What is more, orthogonality of HALO- and SNAP-tag even allows to perform these tasks simultaneously, in order broadly interfere in cellular processes.

2.5. Targeted C-to-U editing with SNAP-CDAR-S

2.5.1. SNAP-CDAR-S is efficient and accurate

While we demonstrated that the SNAP-tag could not only be used for creating a novel C-to-U editor but also in combination with a different protein tag in an orthogonal manner to conduct both A-to-I and C-to-U editing on an endogenous transcript, some issues remained. For some targets mApo1S showed high editing capacity but it was very difficult to deduce a universally valid approach considering codon scope or guideRNA design, rendering it a suboptimal tool. Luckily, Abudayyeh et al. extensively mutated the deaminase domain of ADAR2Q resulting in a C-to-U editor.[191] This mutant in combination with the Cas13 enzyme was termed RESCUE. Given its origin, transcriptome sequencing revealed a substantial number of both, A-to-I and C-to-U off-target editing sites, which prompted Abudayyeh and colleagues to further mutate the enzyme. This led to an enzyme with a reduced number of off-target sites but also decreased on-target editing efficiency, naming it RESCUE-S.[191] In previous work, Cox et al. have shown that fusing the Cas13 enzyme to the deaminase domain of ADAR2Q resulted in an A-to-I editing platform, which however showed substantial short comings when benchmarked with the SNAP-ADAR platform.[182, 190] Therefore, and because this mutated ADAR2 deaminase domain showed remarkable potential, we sought to test if the SNAP-tag would constitute a better option for this enzyme. To this end, Moritz Stoll and Aline Maria Mack as part of her laboratory rotation under my co-supervision created two HEK 293 Flp-In™ T-REx™ cell lines, respectively, stably expressing SNAP-tagged RESCUE (i.e. SNAP-CDAR, by Moritz Stoll) and SNAP-tagged RESCUE-S (i.e. SNAP-CDAR-S, by Aline Maria Mack) deaminase domain, respectively. Design of constructs and conception of cloning and experimental plan was performed by me together with Thorsten Stafforst.

Most intriguingly, we were not able to examine editing capacity of SNAP-CDAR, as its expression proved to be detrimental for the cells. Therefore, all following experiments were performed by me with SNAP-CDAR-S.

For guideRNA design Abudayyeh et al. recommended designing multiple guideRNAs with C-C or C-U mismatches at different positions within the guideRNA/target mRNA hybrid, as editing varied upon that for every target site individually.[191] Therefore, we first investigated and determined the ideal guideRNA design. However, in contrast to the Cas13-based editor, we sought to deduce a generally valid guideRNA design that could be utilized for every transcript and every target site.[191] To this end, we tested three designs differing

in the position of the mismatch C and applied the following nomenclature, we already established for SNAP-ADAR guideRNAs: 5'- X-C-X, with X being the number of nucleotides that bind to the target sites and C constituting the mismatch C. Unless differently stated, all guideRNAs possessed a non-binding extension of three nucleotides at the $(snap)_2$-bearing end to provide extra flexibility (Manuscript 1, Figure 2A & C). Hence, we initially tested 5'- 3-C-18, 10-C-11, and 18-C-3 designs on Q95 on eGFP in a 5'-ACA context that could also be edited by Apo1S. Similar to the observation made for RESCUE-S, SNAP-CDAR-S showed best editing when the mismatch was close to the 5'-end of the guideRNA/mRNA duplex (Manuscript 1, Figure 2C).[191] With a central mismatch, a design similar to that of the SNAP-ADAR-approach, the editing was approximately half of that with the 5'-3-C-18 guideRNA.[182, 268] Placing the mismatch on the 3' end did not yield any editing (Manuscript 1, Figure 2C).

We further assessed SNAP-CDAR-S editing efficiency by targeting endogenously expressed transcripts. On the one hand, we focused on R7 of PPIB, a target that was efficiently edited by Abudayyeh et al.[191] On the other hand, we focused on T52 of GAPDH, a target that was already editable by Apo1S and contained multiple cytidines in close vicinity to each other in the extended site, giving us the possibility to challenge programmability and investigate susceptibility to by-stander editing.

We targeted almost every cytidine ±2-3 nucleotides up- and downstream of T52 target C and examined the surrounding for by-stander editing. Strikingly, cytidines in every examined context except for 5'- CCA showed substantial editing yield while exhibiting background-level by-stander editing (Manuscript 1, Supporting Figure S10). This excellent programmability is in accordance with what has been observed with RESCUE-S indicating that accuracy of the mutated deaminase domain is not reduced by the SNAP-tag.[191] In contrast, activity of the deaminase domain might be impacted more severely. This is indicated by the fact that expressing SNAP-CDAR shows grave cytotoxicity, while expression of its Cas13-based counterpart (i.e. RESCUE) does not.

This was further substantiated by editing of R7 on PPIB, a target that was reported to be edited by RESCUE and to a low degree by RESCUE-S.[191] Here, the guideRNA design 3-C-18 showed editing yields similar to reported yields for the former but about three times higher than those for the latter enzyme (Manuscript 1, Figure 2D).[191] To reiterate, our SNAP-CDAR-S construct contains the deaminase domain of less active RESCUE-S construct.[191] This effect of fusion partner had also been reported before for targeted A-to-I

editing with SNAP-ADARs. There, editing activity was also substantially higher when using SNAP-tag as compared to Cas13 for translocation to target site.[182, 190]

Most intriguingly, when targeting PPIB, the best functioning RESCUE-S guideRNA was longer (30 nucleotides) with the mismatch C placed at position 24 (i.e., position 7 following our nomenclature).[191] Consequently, we compared our initial design (3-C-18) with a design resembling that i.e., 6-C-23. Strikingly, editing with SNAP-CDAR-S was increased with the new design (Manuscript 1. Figure 2D). To investigate if this effect was due to the length of the guideRNA or the positioning of the mismatch C, we employed further designs. On the one hand, we kept the length of 22 nucleotides but shifted the position to 7 (6-C-15) and on the other hand with a guideRNA length of 30 nucleotides we shifted positions to 3, 4, 5, 9, and 11, respectively (Manuscript 1, Figure 2D). Interestingly, when guideRNA length was 22 nucleotides, position 7 of the mismatch only slightly increased editing. In contrast, when increasing size of the guideRNA to 30 nucleotides, editing was severely boosted for mismatch positions 4, 5, and 7 (i.e., 3-C-26, 4-C-25, 6-C-23), confirming moderate design tolerance already observed for RESCUE-S (Manuscript 1, Figure 2D).[191] However, our aim was to deduce a universally valid design that could be used for every target site with minimal to ideally no reduction in editing yield. Therefore, we tested 6-C-23 and 3-C-26 on four targets on endogenous transcripts. We targeted S727F on STAT1 (5'- U**C**U) and on STAT3 (5'- U**C**C), respectively, H36Y (5'- C**C**A) and T41I (5'- A**C**C) on β-catenin (CTNNB1), respectively (Manuscript 1, Supporting Figure S8). Different from editing of PPIB, 6-C-23 outperformed 3-C-26 for every target site. Most intriguingly, STAT1 could only be edited by 6-C-23 (Manuscript 1, Supporting Figure S8). This strongly indicated that 6-C-23 could be considered a universally valid guideRNA design. To further substantiate this, we challenged this design with guideRNA designs reported ideal for RESCUE/RESCUE-S.[191] For this we targeted again S727F on STAT1, H36Y, T41I on CTNNB1, and S21S (5'- C**C**G) on PPIB additionally, with SNAP-CDAR-S. Again, for every target site, 6-C-23 demonstrated superior editing (Manuscript 1, Supporting Figure S9). With this, we were able to demonstrate that we can establish a universally valid $(snap)_2$-guideRNA for targeted C-to-U editing with SNAP-CDAR-S.

To examine if this guideRNA does not lead to undesired by-stander editing, we again targeted the same cluster of cytidines on GAPDH (Manuscript 1, Figure 2F). Here, for all target sites, except for one editing could be enhanced while by-standers remained at background level. Interestingly, S51S (5'- C**C**A) could be elevated above background level

(from 4 to 8 %). In contrast, T52I (5'- A**C**C) was less efficiently edited (41 % vs. 56 %). Editing of V60 (5'- U**C**A) remained at 10 % (Manuscript 1, Figure 2F).

In summary, we successfully transferred the cytidine deaminase domain of RESCUE-S to our SNAP-tag platform and created a tool with high efficiency, maximum programmability, and accuracy at single-nucleotide resolution, even when targets lied among multiple potential substrates. In addition, unlike RESCUE/RESCUE-S, we were able to deduce a universally valid guideRNA design, that comes with only marginal differences in editing yields and does not require screening of multiple designs.[191]

2.5.2. SNAP-CDAR-S editing is tunable and superior to RESCUE-S

As guideRNA design could be optimized leading to editing yields higher than reported yields for the Cas13-based cytidine deaminase[191], we sought to examine further aspects of our system. For A-to-I editing with SNAP-ADAR it had already been shown that editing can be modulated by time and guideRNA amount.[182] Therefore, we investigated these aspects.

We first tested different amounts of $(snap)_2$-guideRNA targeting PPIB R7C over the course of a regular editing experiment, i.e., 48 h (Manuscript 1, Figure 2E). When applying 2.5 pmol per 8 x 10^4 cell we can almost see a maximum editing yield, as it only slightly increases when applying double the amount of guideRNA. This suggests that at 5 pmol our system is fully saturated. This could also be seen when experiment endpoint was reduced to 24 h (Manuscript 1, Figure 2E). There, virtually no increase in editing was observed when doubling the applied guideRNA amount. However, for almost all guideRNA amounts, editing dropped when length of experiments was shortened, suggesting that cytidine deamination with SNAP-CDAR-S seems as slow a process, as it is with mApo1S (Manuscript 1, Figure 2E). In fact, considerably slow deamination was also observed for all systems that perform C-to-U editing, even when utilizing native cytidine deaminases such as APOBECs.[191, 194, 195] In contrast to that, A-to-I editing with SNAP-ADARs almost peaked 3 h post guideRNA application.[182]

Next, we pursued a benchmark with RESCUE-S. A remote comparison would not be conclusive, as these approaches do not only differ in the protein moiety promoting translocation to target site but also in delivery of the editase itself. As already mentioned, our editase is stably integrated into HEK 293 Flp-In™ T-REx™ cells with a doxycycline-induced expression. The guideRNAs are then as such transfected into the cells, 24 h post-

induction. In contrast, Abudayyeh et al. transfect all components, that is editase, guideRNA, and where applicable the reporter as plasmid DNA.[191] Therefore, to generate better comparability, Moritz Stoll created another HEK 293 Flp-In™ T-REx™ cell line for doxycycline-induced RESCUE-S expression. Like this, only guideRNA and where applicable reporter pDNA were transfected. Design of constructs and conception of cloning and experimental plan were performed by me together with Thorsten Stafforst. Editing experiments were performed by me.

For the benchmark, we chose sites that were already reported to be susceptible to C-to-U editing by Abudayyeh et al.[191] We further included the T52 site of GAPDH, as one of our own (Manuscript 1. Figure 3A). We applied the best working guideRNA design by Abudayyeh et al. for the reported sites. Briefly, SNAP-CDAR-S outperformed RESCUE-S in every single instance. Within the targeted sites, we showed that also functional sites can be altered such as phosphoserines in STAT1 and STAT3 or phosphothreonine of CTNNB1, all of which to a higher degree than RESCUE-S (Manuscript 1, Figure 3A). Furthermore, we targeted two arginines in two different sequence contexts in different transcripts. In PPIB the 5'-ACG context was very efficiently edited by SNAP-CDAR-S, about five-fold better than with RESCUE-S. The other arginine target was within a disease-associated context. The target transcript codes for a protein that exists in three different isoforms in the general population. The isforms, APOE2-4, are formed by polymorphisms of residues 112 and 158.[269] The rarest form APOE2 bears two cysteines at those sites. In contrast, APOE4 bears two arginines and constitutes a risk factor for Alzheimer's disease. APOE3 is a mix bearing a cysteine at 112 and an arginine at 158 and is the most common polymorphism. While APOE3 and 4 are associated with Alzheimer's disease, APOE2 is considered protective.[269] Therefore, editing of that target is potentially of clinical interest. Besides disease-relevance of this target, the edited C lies in a 5'-GCG context. Guanine is a severely unpreferred 5' nearest neighbor for all ADAR deaminase domains.[182] Luckily, this site showed some degree of editing for both, SNAP-CDAR-S, and RESCUE-S. Again, SNAP-CDAR-S was superior, yielding even higher editing than reported for the more active RESCUE system (Manuscript 1, Figure 3A).[191] Unfortunately, RESCUE-S was not utilized by Abudayyeh et al. for this target.[191] In contrast, R112 did not exhibit any susceptibility to editing. A possible reason for this might be found in the extended very GC-rich sequence context. In particular for R112 this resulted in a predicted secondary structure of the target site that seems substantially stronger than for R158 (Figure 16).

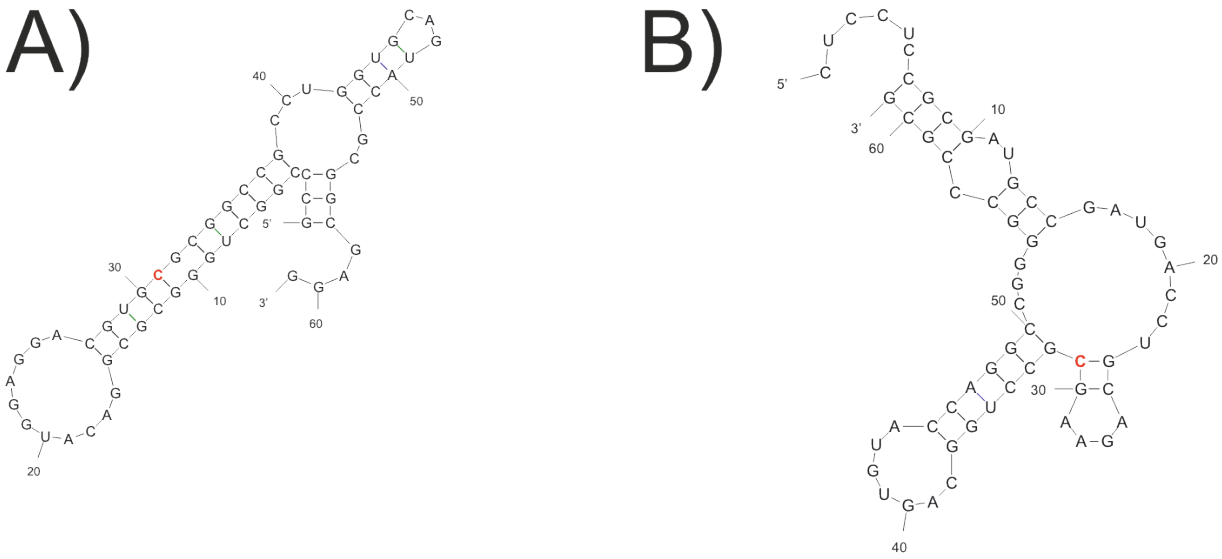


Figure 16: ***Predicted secondary structures of APOE target sites.*** ***A)*** *Predicted secondary structure of APOE target site for R112,* ***B)*** *Predicted secondary structure of APOE target R158. Secondary structure predictions were performed with mfold.*[270]

This could explain the discrepancy. Fortunately, when editing R158, a lower risk (APOE4→APOE3) or even a protective (APOE3→APOE2) genotype can be created.[269]
Finally, we challenged all possible 5'-CAN codons. As explained above, guanine's amino group can cause spatial perturbation (section 1.1.5, Figure 2C).[142] Similar to A-to-I editing with SNAP-ADARs, also here the full potential of our system is evident, as we substituted the guanine opposite of the 5'-C with inosine (Manuscript 1, Figure 2H). When a guanosine was present, CCU, CCG, and CCA only showed about 10% editing, respectively and CCC showed about 30%. However, when inosine was incorporated editing was dramatically elevated, of about 3-fold to 30-35 % for CCU, CCG, and CCA, respectively and more than double to about 70 % for CCC (Manuscript 1, Figure 2H). In stark contrast, none of the tested codons could be edited by RESCUE-S (5'-CCC was not tested), as already described by Abudayyeh et al (Manuscript 1, Figure 3A).[191]
Next, we analyzed effect of editing on a protein level. For this, we focused on two assays. Firstly, we targeted phosphorylation site T41 within CTNNB1, which is known to be involved in its degradation upon phosphorylation.[271] Subsequently, induced T41I mutation by C-to-U editing would lead to a stable CTNNB1, which then would accumulate in the nucleus where it would trigger T cell factor/lymphoid enhancer factor (TCF/LEF) induced gene expression.[272] We used a reporter plasmid encoding FireFly luciferase, with TCF/LEF responsive elements. By extension, these elements are CTNNB1-responsive. For simplification, we will refer to it as such. Consequently, editing of CTNNB1 T41 would

lead to increased FireFly luciferase expression. We determined both, editing on RNA level, and increase of FireFly luciferase signal for both systems, respectively (Manuscript 1, Figure 3B, Supporting Figure S11). The results confirmed what we had seen before. SNAP-CDAR-S by far exceeded editing potency of RESCUE-S. While editing on RNA level was as expected substantially higher for SNAP-CDAR-S than for RESCUE-S, the increase in signal was also phenomenal. Editing of the T41 with SNAP-CDAR-S increased the FireFly signal about eight-fold to the unedited sample. In stark contrast, the 7% editing yield by RESUCE-S increased FireFly signal only by about a 1.7-fold (Manuscript 1, Figure 3B, Supporting Figure S11). Unfortunately, we cannot assess, if our results are in accordance with what was observed by Abudayyeh et al, as they have not executed this assay with RESCUE-S.[191]

Secondly, we focused on abolishing another phosphorylation site, that is serine 727 of STAT3. As protein reduction by RNA editing can potentially be difficult, due to protein stability, we sought to keep editing stable over multiple days. For this, Lipofectamine 2000 proved to be detrimental for cell survival, as after the second transfection all cells degraded. Consequently, we used RNAiMAX in a similar fashion, which is why we could not perform this assay with RESCUE-S. We transfected 5 pmol of $(snap)_2$-guideRNA targeting STAT3 S727 every 48 hours over eight days. At the endpoint, we took one part of the cells for RNA isolation and determination of editing yield on RNA level and the rest for Western Blot. Editing on transcript level was stable over eight days. Appropriate controls demonstrated the accuracy of our system, as even after that prolonged amount of time editing was at background level (Manuscript 1, Figure 3C). This editing could then also be seen on protein level, as a clearly visible reduction of antibody signal for S727 was observed for samples exhibiting editing on RNA level (Manuscript 1, Figure 3C).

By extensively mutating ADAR2's deaminase domain, Abudayyeh and colleagues undoubtedly created an effective programmable cytidine deaminase with a broad codon scope.[191] However, same as for REPAIR, Cas13 was not the most ideal fusion partner for a very potent RNA base deamination tool.[182, 190] In contrast, here we show that with the help of the SNAP-tag, we created a highly potent tool for site-directed C-to-U RNA editing with maximum programmability. What is more, by utilizing chemically modified guideRNAs we were able to substantially increase editing of difficult 5'-CCN codons. Most importantly, this efficiency in editing we were even able to demonstrate on a protein level. While other recently developed tools based APOBEC-3A or further developed RESCUE variants demonstrate high activity, they still come with some downsides. For one, CURE

and REWIRE have an inherently restricted codon scope owing to utility of APOBEC-3A as the catalytically active moiety. What is more, DNA-editing activity of APOBEC-3A poses another issue with regards to safety.[195, 219] In contrast, eRESCUE, which utilizes dPspCas13b instead of dRanCas13b, exhibits very high editing capacity. In addition, Li and colleagues benchmarked eRESCUE with RESCUE and demonstrated eRESCUE's superiority, contradicting the report by Abudayyeh et al., who demonstrated that both Cas13b orthologs possess equal activity for targeted C-to-U editing.[191, 213] Here, a benchmark with our system might be essential to assess, if dPspCas13b-RESCUE-S could indeed yield editing levels rivaling SNAP-CDAR-S. Unfortunately, despite mentioning and terming dPspCas13b-RESCUE-S variant as "eRESCUE-S", the authors did not provide editing data, which is why an adequate prediction is not possible.[213] Lastly, the most recently developed tool, split-RESCUE, elicited very low levels of editing (below 6 %) on endogenous targets.[197]

Nevertheless, with regard to SNAP-ADARs, editing levels of SNAP-CDAR-S are rather low. Consequently, there is plenty room for improvement. In 2021 Malik and colleagues reported a pH-dependency of ADAR's editing activity. More accurately, under acidic conditions the strand invading glutamate is fully protonated and can therefore form stronger H-bond with the mismatch C. In contrast, under physiological conditions a glutamine on that site is already fully protonated and constantly forms H-bond with the mismatch C, thus explaining why the EQ-variants of both ADARs exhibit hyperactivity.[147] In the same year Doherty et al. then demonstrated that by using a cytidine analog that itself can form an H-bond with the glutamate, called Benner's base, editing with endogenous wildtype ADARs can be severely boosted.[229] Subsequently, by mutating Q488 of SNAP-CDAR (expression was detrimental to cells) to E488 we could potentially create a tool with a reduced editing activity. However, by incorporating Benner's base in our guideRNA, reduction of activity could be compensated. What is more, this reconstitution of activity would be strictly guideRNA-dependent and would thus potentially not demonstrate the same lethality to the cell as SNAP-CDAR. Similar to Malik et al. in 2021, this could be tested by using a cell culture medium with reduced buffer to acidify the medium over the course of an editing experiment and assess if CDAR-Q488E activity would be equally pH-dependent as the ADAR deaminase domains.[147]

In summary, we created a very potent and capable tool, which still exhibits potential to grow. Unlike tools dependent on encoded guideRNAs, we are able to assist that growth by utilizing chemically modified bases in combination with protein engineering.

2.5.3. SNAP-CDAR-S transcriptome-wide editing profile

The glaring superiority of our approach over the Cas13-based approach was evident by a high on-target efficiency with maximum programmability. While endogenous targets were substantially more efficiently edited by SNAP-CDAR-S than RESCUE-S, we hypothesized that it was based on a higher intensity of the editase, which in turn leaves the possibility for a similarly higher transcriptome-wide off-target profile. Therefore, we performed transcriptome sequencing for the two systems. For this, the transcriptome-wide off-target profile of the editase in absence and presence of a guideRNA targeting PPIB R7C, respectively was compared to the transcriptome-wide editing profile of HEK 293 Flp-In™ T-REx™ cells stably expressing an empty transgene cassette. As efficient editing, we only considered sites that were over 5 % more edited than in empty HEK 293 Flp-In™ T-REx™ cells. As the cytidine deaminase is derived from the hyperactive variant of ADAR2 deaminase domain, we considered both, A to I, and C to U editing events.[191]

For both systems we recorded moderate levels of A to I off-target editing and quite low levels of C to U off-target editing (Manuscript 1 Figure 4B & C). By far most of the off-targets of both kinds could be traced back to the over-expression of the editase and not guideRNA fidelity (Manuscript 1, Figure 4C & G). This was also observed for SNAP-ADARs.[182] While the overall number of total sites was comparable among the two systems, intensity was higher for SNAP-CDAR-S, which confirms our results from the benchmark of our systems on endogenous targets (Manuscript 1, Figure 4D).

Intriguingly, for RESCUE-S we recorded more guideRNA dependent off-targets than for SNAP-CDAR-S. In fact, there were about three times more gRNA-dependent C-to-U off-targets (Manuscript 1, Figure 4G). This could be accredited to the chemical modifications utilized in our guideRNAs, as it is known that editing can be extinguished by 2'OMe modifications (our guideRNAs bear 2'OMe modifications at every nucleotide except the mismatch gap).[168] In contrast, plasmid-borne RESCUE-S guideRNAs only contain natural nucleotides and ADARs were reported to arbitrarily edit all adenines within an RNA duplex.[14, 15, 18]

Other tools based on fusion proteins exhibited varying degrees of off-targets. On the one hand, CURE, the Cas13-APOBEC-3A fusion protein exhibited more than twice as many C-to-U off-target sites as APOBEC-3A alone (7416 vs. 3269). However, this could be strongly reduced by fusion of APOBEC-3A to dCasRx. While largely retaining on-target editing, the authors reported that this CURE-X induced fewer off-target sites than RESCUE-S (380 vs. 611 sites), most of which showed below 25 % editing.[195] On the other

hand, for CU-REWIRE, which also utilizes APOBEC-3A as a deaminase only 711 transcriptome-wide C-to-U off-target sites were reported, which however could be slightly reduced to 614 sites, while simultaneously strongly increasing on-target editing by increasing the PUF domain to bind ten nucleotides (CU-REWIRE3.0).[219] In contrast to those tools, eRESCUE elicited a very large number of both A-to-I, and C-to-U off-target editings.[213] Similar to our tool, eRESCUE showed substantially higher editing activity than its ancestral counterpart (here: RESCUE). However, unlike SNAP-CDAR-S, eRESCUE exhibited extremely low accuracy, eliciting both, A-to-I, and C-to-U editing at over ten-times more sites than RESCUE.[213] Most intriguingly, both tools exhibited varying numbers of off-targets dependent on the targeted transcript. While slight variation in the number of both A-to-I and C-to-U editing was found for eRESCUE, more than double the number for A-to-I and more than four times the number for C-to-U editing sites depending on the targeted transcript was found for RESCUE.[213]

Finally, in 2022 Katrekar and colleagues reported a high-potential approach for severely reducing transcriptome-wide off-target sites. In this approach the deaminase domain of ADAR2 and RESCUE, that is CDAR, is split in two fragments. Each fragment is fused to either MS2 coat protein or the λ-N peptide.[197] The guideRNA bearing the corresponding hairpins tether the fusion proteins for assembly at the target site. While editing yields were rather low (below 40 % for A-to-I and below 6 % for C-to-U), transcriptome-wide off-target sites were at background level.[197] As described above, there is still room for improvement of our system and one way would be protein engineering of a CDAR deaminase domain with higher activity. Unfortunately, this could lead to an increase in off-target editing sites. In accordance with the approach of Katrekar et al.[197], this could be mitigated by splitting the CDAR domain and fuse either part to either SNAP- or HALO-tag. A bi-functional guideRNA bearing both of the according substrates would mediate translocation and assembly at the target site. This way even the CDAR-domain, which was detrimental for the cell when fused to the SNAP-tag, could be utilized.

In addition, Stroppel et al. have shown in 2021 that by fusing the SNAP-tag to gibberellic acid insensitive and the ADAR deaminase domain to gibberellin insensitive dwarf 1A, two plant proteins, successful recruitment with relatively high editing was mediated upon application of gibberellic acid to the cell culture medium in a dose-dependent manner. This could also be a viable option.

In summary, with moderate levels of transcriptome-wide off-target editing and decent editing rates with minimal restrictions in editable codons, which could even be transferred to the protein level SNAP-CDAR-S is a high-value tool with yet a lot of potential for development. The combination of protein engineering and employment of chemically modified nucleotides does not just provide means to develop a tool that can show even higher editing levels but simultaneously provides the possibility to control by-standers and thus secure accuracy of the system.

3. Conclusion and future prospects

The targeted RNA-editing field has grown very quickly over the past few years, comprising a wide and manifold range of different techniques. Some recruit endogenously expressed ADARs constituting a robust and very promising platform for future therapeutic approaches that due to the elicited transient effect are not subject to serious ethical reservations raised in the surge and recent advance of genome editing technologies.[273] In contrast to that, researchers have established a variety of fusion proteins for targeted A-to-I editing. One of those systems strongly relies on chemistry. Here, recruitment of editing moiety is provided by a stoichiometric coupling of the SNAP-tagged editor to the substrate-bearing guideRNA. This approach further relies on chemically synthesized guideRNAs, which leaves the possibility of implementing known modifications to increase stability and perform other tasks.[180-182]

In this work we successfully optimized guideRNA design to the point that we determined minimal length still largely preserving editing capacity. What is more, we developed a modification pattern that on the one hand increases stability in extremely hostile environment such as lysosome lysate and on the other hand provides the capacity to be taken up by the cells when supplied in the culture medium without any complexing agent.

In addition, we further developed the SNAP-ADAR platform by implementing different self-labeling enzymes with their respective substrate to orthogonally and simultaneously perform both A-to-I and C-to-U editing. As editing yields for the latter were erratic and rather low, we decided to implement the SNAP-tag for targeted C-to-U editing with the mutated ADAR2 deaminase domain. We could show that not only is our SNAP-CDAR-S effective, but also that the chemical modifications of the guideRNAs allowed us to substantially edit difficult 5'- CCN codons. Furthermore, this extensive editing capacity did not negatively impact accuracy, as it showed a similar transcriptome-wide editing profile as its lower performing Cas13-based counterpart RESCUE-S.

However, while SNAP-CDAR-S constitutes a very promising tool, editing yields for some targets remained very low. In addition, editing of the best editable sites PPIB R7C and STAT3 S727F could not pass 50 %. Compared to established A-to-I editors that yield editing of up to 90 %, there is much room for improvement.[182] By introduction of further mutations, one could further attempt to increase editing. However, considering that SNAP-CDAR proved detrimental for the expressing cell lines within an entire editing experiment

time frame, an editing activity between that of SNAP-CDAR and SNAP-CDAR-S should be aspired.

Nevertheless, with this strong cytidine editor we can now seek to further develop and refine orthogonal mRNA editing. Combining HALO- and SNAP-tag with ADAR and CDAR-S deaminase domains would constitute a strong a viable expansion to the mRNA editing toolbox.

On the other hand, another possibility would constitute construction of an editor that consists of an ADAR deaminase domain connected with the CDAR-S deaminase domain via a SNAP-tag or HALO-tag. This would provide the following advantages. For one, whether adenosines or cytidines are targeted could be specified by applied guideRNA (A:C mismatch for adenosine targeting and C:C mismatch for cytidine targeting). Secondly, the very small size (estimated size of approximately 3.1-3.4 kb) would make it possible to create an AAV, by which potentially primary cells can be targeted for editing. This would further allow for application in animal models. In a recent publication authors showed discovery and implementation of a compact Cas13, which is substantially smaller than the other dPspCas13b/dRanCas13b for targeted A-to-I and C-to-U editing. Due to this, REPAIR could successfully be encoded with an AAV, albeit showing very low editing when applied in this form. Still, this small size does not allow for creation of an AAV encoding for both REPAIR and RESCUE.[210] Here, we could provide a viable alternative. What is more, implementation of chemical modifications would allow for targeting of specific tissues or organs in animal models.

Taken together, the chemical modifications of the guideRNAs and the ready implementation of the SNAP-tag to all editing enzymes provided great potential that we could largely tap to lay the pavement for exciting novel application techniques and the possibility for multiplexing of editing systems that go beyond the scope of A-to-I and C-to-U editing.

4. References

1. Petrenko, V., Sinturel, F., Riezman, H. & Dibner, C. Lipid metabolism around the body clocks. *Prog Lipid Res* **91**, 101235 (2023).
2. Scott, J. et al. RNA editing: a novel mechanism for regulating lipid transport from the intestine. *Gut* **30 Spec No**, 35-43 (1989).
3. Young, S.G. Recent progress in understanding apolipoprotein B. *Circulation* **82**, 1574-1594 (1990).
4. Blue, M.L., Protter, A.A. & Williams, D.L. Biosynthesis of apolipoprotein B in rooster kidney, intestine, and liver. *J Biol Chem* **255**, 10048-10051 (1980).
5. Teng, B., Verp, M., Salomon, J. & Davidson, N.O. Apolipoprotein B messenger RNA editing is developmentally regulated and widely expressed in human tissues. *J Biol Chem* **265**, 20616-20620 (1990).
6. Glickman, R.M., Rogers, M. & Glickman, J.N. Apolipoprotein B synthesis by human liver and intestine in vitro. *Proc Natl Acad Sci U S A* **83**, 5296-5300 (1986).
7. Chen, S.H. et al. Apolipoprotein B-48 is the product of a messenger RNA with an organ-specific in-frame stop codon. *Science* **238**, 363-366 (1987).
8. Powell, L.M. et al. A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine. *Cell* **50**, 831-840 (1987).
9. Hospattankar, A.V., Higuchi, K., Law, S.W., Meglin, N. & Brewer, H.B., Jr. Identification of a novel in-frame translational stop codon in human intestine apoB mRNA. *Biochem Biophys Res Commun* **148**, 279-285 (1987).
10. Davidson, N.O., Powell, L.M., Wallis, S.C. & Scott, J. Thyroid hormone modulates the introduction of a stop codon in rat liver apolipoprotein B messenger RNA. *J Biol Chem* **263**, 13482-13485 (1988).
11. Higuchi, K. et al. Human apolipoprotein B (apoB) mRNA: identification of two distinct apoB mRNAs, an mRNA with the apoB-100 sequence and an apoB mRNA containing a premature in-frame translational stop codon, in both liver and intestine. *Proc Natl Acad Sci U S A* **85**, 1772-1776 (1988).
12. Tennyson, G.E., Sabatos, C.A., Eggerman, T.L. & Brewer, H.B., Jr. Characterization of single base substitutions in edited apolipoprotein B transcripts. *Nucleic Acids Res* **17**, 691-698 (1989).
13. Bostrom, K. et al. Apolipoprotein B48 RNA editing in chimeric apolipoprotein EB mRNA. *J Biol Chem* **264**, 15701-15708 (1989).
14. Bass, B.L. & Weintraub, H. A developmentally regulated activity that unwinds RNA duplexes. *Cell* **48**, 607-613 (1987).
15. Rebagliati, M.R. & Melton, D.A. Antisense RNA injections in fertilized frog eggs reveal an RNA duplex unwinding activity. *Cell* **48**, 599-605 (1987).
16. Wagner, R.W. & Nishikura, K. Cell cycle expression of RNA duplex unwindase activity in mammalian cells. *Mol Cell Biol* **8**, 770-777 (1988).
17. Wagner, R.W. et al. Double-stranded RNA unwinding and modifying activity is detected ubiquitously in primary tissues and cell lines. *Mol Cell Biol* **10**, 5586-5590 (1990).
18. Bass, B.L. & Weintraub, H. An unwinding activity that covalently modifies its double-stranded RNA substrate. *Cell* **55**, 1089-1098 (1988).
19. Wagner, R.W., Smith, J.E., Cooperman, B.S. & Nishikura, K. A double-stranded RNA unwinding activity introduces structural alterations by means of adenosine to inosine conversions in mammalian cells and Xenopus eggs. *Proc Natl Acad Sci U S A* **86**, 2647-2651 (1989).

20. Polson, A.G., Crain, P.F., Pomerantz, S.C., McCloskey, J.A. & Bass, B.L. The mechanism of adenosine to inosine conversion by the double-stranded RNA unwinding/modifying activity: a high-performance liquid chromatography-mass spectrometry analysis. *Biochemistry* **30**, 11507-11514 (1991).
21. Sommer, B., Kohler, M., Sprengel, R. & Seeburg, P.H. RNA editing in brain controls a determinant of ion flow in glutamate-gated channels. *Cell* **67**, 11-19 (1991).
22. Bass, B.L. et al. A standardized nomenclature for adenosine deaminases that act on RNA. *RNA* **3**, 947-949 (1997).
23. Chen, C.X. et al. A third member of the RNA-specific adenosine deaminase gene family, ADAR3, contains both single- and double-stranded RNA binding domains. *RNA* **6**, 755-767 (2000).
24. Nishikura, K. Functions and regulation of RNA editing by ADAR deaminases. *Annu Rev Biochem* **79**, 321-349 (2010).
25. Nishikura, K. A-to-I editing of coding and non-coding RNAs by ADARs. *Nat Rev Mol Cell Biol* **17**, 83-96 (2016).
26. Kim, U. et al. Purification and characterization of double-stranded RNA adenosine deaminase from bovine nuclear extracts. *J Biol Chem* **269**, 13480-13489 (1994).
27. Hough, R.F. & Bass, B.L. Purification of the Xenopus laevis double-stranded RNA adenosine deaminase. *J Biol Chem* **269**, 9933-9939 (1994).
28. Herbert, A. et al. A Z-DNA binding domain present in the human editing enzyme, double-stranded RNA adenosine deaminase. *Proc Natl Acad Sci U S A* **94**, 8421-8426 (1997).
29. Herbert, A. & Rich, A. The role of binding domains for dsRNA and Z-DNA in the in vivo editing of minimal substrates by ADAR1. *Proc Natl Acad Sci U S A* **98**, 12132-12137 (2001).
30. Kim, U., Wang, Y., Sanford, T., Zeng, Y. & Nishikura, K. Molecular cloning of cDNA for double-stranded RNA adenosine deaminase, a candidate enzyme for nuclear RNA editing. *Proc Natl Acad Sci U S A* **91**, 11457-11461 (1994).
31. O'Connell, M.A. et al. Cloning of cDNAs encoding mammalian double-stranded RNA-specific adenosine deaminase. *Mol Cell Biol* **15**, 1389-1397 (1995).
32. O'Connell, M.A. & Keller, W. Purification and properties of double-stranded RNA-specific adenosine deaminase from calf thymus. *Proc Natl Acad Sci U S A* **91**, 10596-10600 (1994).
33. Melcher, T. et al. A mammalian RNA editing enzyme. *Nature* **379**, 460-464 (1996).
34. Marcucci, R. et al. Pin1 and WWP2 regulate GluR2 Q/R site RNA editing by ADAR2 with opposing effects. *EMBO J* **30**, 4211-4222 (2011).
35. Melcher, T. et al. RED2, a brain-specific member of the RNA-specific adenosine deaminase family. *J Biol Chem* **271**, 31795-31798 (1996).
36. Driscoll, D.M. & Casanova, E. Characterization of the apolipoprotein B mRNA editing activity in enterocyte extracts. *J Biol Chem* **265**, 21401-21403 (1990).
37. Lau, P.P., Chen, S.H., Wang, J.C. & Chan, L. A 40 kilodalton rat liver nuclear protein binds specifically to apolipoprotein B mRNA around the RNA editing site. *Nucleic Acids Res* **18**, 5817-5821 (1990).
38. Backus, J.W. & Smith, H.C. Apolipoprotein B mRNA sequences 3' of the editing site are necessary and sufficient for editing and editosome assembly. *Nucleic Acids Res* **19**, 6781-6786 (1991).
39. Greeve, J., Navaratnam, N. & Scott, J. Characterization of the apolipoprotein B mRNA editing enzyme: no similarity to the proposed mechanism of RNA editing in kinetoplastid protozoa. *Nucleic Acids Res* **19**, 3569-3576 (1991).

40. Smith, H.C. et al. In vitro apolipoprotein B mRNA editing: identification of a 27S editing complex. *Proc Natl Acad Sci U S A* **88**, 1489-1493 (1991).
41. Harris, S.G. et al. Extract-specific heterogeneity in high-order complexes containing apolipoprotein B mRNA editing activity and RNA-binding proteins. *J Biol Chem* **268**, 7382-7392 (1993).
42. Giannoni, F. et al. Complementation of apolipoprotein B mRNA editing by human liver accompanied by secretion of apolipoprotein B48. *J Biol Chem* **269**, 5932-5936 (1994).
43. Driscoll, D.M. & Zhang, Q. Expression and characterization of p27, the catalytic subunit of the apolipoprotein B mRNA editing enzyme. *J Biol Chem* **269**, 19843-19847 (1994).
44. Schock, D. et al. An auxiliary factor containing a 240-kDa protein complex is involved in apolipoprotein B RNA editing. *Proc Natl Acad Sci U S A* **93**, 1097-1102 (1996).
45. Navaratnam, N., Shah, R., Patel, D., Fay, V. & Scott, J. Apolipoprotein B mRNA editing is associated with UV crosslinking of proteins to the editing site. *Proc Natl Acad Sci U S A* **90**, 222-226 (1993).
46. Teng, B., Burant, C.F. & Davidson, N.O. Molecular cloning of an apolipoprotein B messenger RNA editing protein. *Science* **260**, 1816-1819 (1993).
47. Navaratnam, N. et al. The p27 catalytic subunit of the apolipoprotein B mRNA editing enzyme is a cytidine deaminase. *J Biol Chem* **268**, 20709-20712 (1993).
48. MacGinnitie, A.J., Anant, S. & Davidson, N.O. Mutagenesis of apobec-1, the catalytic subunit of the mammalian apolipoprotein B mRNA editing enzyme, reveals distinct domains that mediate cytosine nucleoside deaminase, RNA binding, and RNA editing activity. *J Biol Chem* **270**, 14768-14775 (1995).
49. Yamanaka, S., Poksay, K.S., Balestra, M.E., Zeng, G.Q. & Innerarity, T.L. Cloning and mutagenesis of the rabbit ApoB mRNA editing protein. A zinc motif is essential for catalytic activity, and noncatalytic auxiliary factor(s) of the editing complex are widely distributed. *J Biol Chem* **269**, 21725-21734 (1994).
50. Lau, P.P., Zhu, H.J., Baldini, A., Charnsangavej, C. & Chan, L. Dimeric structure of a human apolipoprotein B mRNA editing protein and cloning and chromosomal localization of its gene. *Proc Natl Acad Sci U S A* **91**, 8522-8526 (1994).
51. Salter, J.D., Bennett, R.P. & Smith, H.C. The APOBEC Protein Family: United by Structure, Divergent in Function. *Trends Biochem Sci* **41**, 578-594 (2016).
52. Smith, H.C. RNA binding to APOBEC deaminases; Not simply a substrate for C to U editing. *RNA Biol* **14**, 1153-1165 (2017).
53. Harris, R.S. & Dudley, J.P. APOBECs and virus restriction. *Virology* **479-480**, 131-145 (2015).
54. Lerner, T., Papavasiliou, F.N. & Pecori, R. RNA Editors, Cofactors, and mRNA Targets: An Overview of the C-to-U RNA Editing Machinery and Its Implication in Human Disease. *Genes (Basel)* **10** (2018).
55. Stavrou, S. & Ross, S.R. APOBEC3 Proteins in Viral Immunity. *J Immunol* **195**, 4565-4570 (2015).
56. Pecori, R., Di Giorgio, S., Paulo Lorenzo, J. & Nina Papavasiliou, F. Functions and consequences of AID/APOBEC-mediated DNA and RNA deamination. *Nat Rev Genet* **23**, 505-518 (2022).
57. Dickerson, S.K., Market, E., Besmer, E. & Papavasiliou, F.N. AID mediates hypermutation by deaminating single stranded DNA. *J Exp Med* **197**, 1291-1296 (2003).

58. Shen, H.M. & Storb, U. Activation-induced cytidine deaminase (AID) can target both DNA strands when the DNA is supercoiled. *Proc Natl Acad Sci U S A* **101**, 12997-13002 (2004).
59. Sharma, S., Patnaik, S.K., Taggart, R.T. & Baysal, B.E. The double-domain cytidine deaminase APOBEC3G is a cellular site-specific RNA editing enzyme. *Sci Rep* **6**, 39100 (2016).
60. Sharma, S. & Baysal, B.E. Stem-loop structure preference for site-specific RNA editing by APOBEC3A and APOBEC3G. *PeerJ* **5**, e4136 (2017).
61. Sharma, S., Patnaik, S.K., Kemer, Z. & Baysal, B.E. Transient overexpression of exogenous APOBEC3A causes C-to-U RNA editing of thousands of genes. *RNA Biol* **14**, 603-610 (2017).
62. Sharma, S. et al. APOBEC3A cytidine deaminase induces RNA editing in monocytes and macrophages. *Nat Commun* **6**, 6881 (2015).
63. Greeve, J., Altkemper, I., Dieterich, J.H., Greten, H. & Windler, E. Apolipoprotein B mRNA editing in 12 different mammalian species: hepatic expression is reflected in low concentrations of apoB-containing plasma lipoproteins. *J Lipid Res* **34**, 1367-1383 (1993).
64. Hadjiagapiou, C., Giannoni, F., Funahashi, T., Skarosi, S.F. & Davidson, N.O. Molecular cloning of a human small intestinal apolipoprotein B mRNA editing protein. *Nucleic Acids Res* **22**, 1874-1879 (1994).
65. Davidson, N.O. et al. Proposed nomenclature for the catalytic subunit of the mammalian apolipoprotein B mRNA editing enzyme: APOBEC-1. *RNA* **1**, 3 (1995).
66. Fujino, T., Navaratnam, N., Jarmuz, A., von Haeseler, A. & Scott, J. C-->U editing of apolipoprotein B mRNA in marsupials: identification and characterisation of APOBEC-1 from the American opossum Monodelphus domestica. *Nucleic Acids Res* **27**, 2662-2671 (1999).
67. Hirano, K. et al. Targeted disruption of the mouse apobec-1 gene abolishes apolipoprotein B mRNA editing and eliminates apolipoprotein B48. *J Biol Chem* **271**, 9887-9890 (1996).
68. Morrison, J.R. et al. Apolipoprotein B RNA editing enzyme-deficient mice are viable despite alterations in lipoprotein metabolism. *Proc Natl Acad Sci U S A* **93**, 7154-7159 (1996).
69. Nakamuta, M. et al. Complete phenotypic characterization of apobec-1 knockout mice with a wild-type genetic background and a human apolipoprotein B transgenic background, and restoration of apolipoprotein B mRNA editing by somatic gene transfer of Apobec-1. *J Biol Chem* **271**, 25981-25988 (1996).
70. Cole, D.C. et al. Loss of APOBEC1 RNA-editing function in microglia exacerbates age-related CNS pathophysiology. *Proc Natl Acad Sci U S A* **114**, 13272-13277 (2017).
71. Lellek, H. et al. Purification and molecular cloning of a novel essential component of the apolipoprotein B mRNA editing enzyme-complex. *J Biol Chem* **275**, 19848-19856 (2000).
72. Mehta, A., Kinter, M.T., Sherman, N.E. & Driscoll, D.M. Molecular cloning of apobec-1 complementation factor, a novel RNA-binding protein involved in the editing of apolipoprotein B mRNA. *Mol Cell Biol* **20**, 1846-1854 (2000).
73. Mehta, A., Banerjee, S. & Driscoll, D.M. Apobec-1 interacts with a 65-kDa complementing protein to edit apolipoprotein-B mRNA in vitro. *J Biol Chem* **271**, 28294-28299 (1996).

74. Mehta, A. & Driscoll, D.M. A sequence-specific RNA-binding protein complements apobec-1 To edit apolipoprotein B mRNA. *Mol Cell Biol* **18**, 4426-4432 (1998).
75. Blanc, V., Henderson, J.O., Kennedy, S. & Davidson, N.O. Mutagenesis of apobec-1 complementation factor reveals distinct domains that modulate RNA binding, protein-protein interaction with apobec-1, and complementation of C to U RNA-editing activity. *J Biol Chem* **276**, 46386-46393 (2001).
76. Henderson, J.O., Blanc, V. & Davidson, N.O. Isolation, characterization and developmental regulation of the human apobec-1 complementation factor (ACF) gene. *Biochim Biophys Acta* **1522**, 22-30 (2001).
77. Lau, P.P., Zhu, H.J., Nakamuta, M. & Chan, L. Cloning of an Apobec-1-binding protein that also interacts with apolipoprotein B mRNA and evidence for its involvement in RNA editing. *J Biol Chem* **272**, 1452-1455 (1997).
78. Anant, S. et al. 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).
79. Blanc, V. et al. Identification of GRY-RBP as an apolipoprotein B RNA-binding protein that interacts with both apobec-1 and apobec-1 complementation factor to modulate C to U editing. *J Biol Chem* **276**, 10272-10283 (2001).
80. Lau, P.P., Chang, B.H. & Chan, L. Two-hybrid cloning identifies an RNA-binding protein, GRY-RBP, as a component of apobec-1 editosome. *Biochem Biophys Res Commun* **282**, 977-983 (2001).
81. Lau, P.P. et al. A DnaJ protein, apobec-1-binding protein-2, modulates apolipoprotein B mRNA editing. *J Biol Chem* **276**, 46445-46452 (2001).
82. Lau, P.P. & Chan, L. Involvement of a chaperone regulator, Bcl2-associated athanogene-4, in apolipoprotein B mRNA editing. *J Biol Chem* **278**, 52988-52996 (2003).
83. Greeve, J., Lellek, H., Rautenberg, P. & Greten, H. Inhibition of the apolipoprotein B mRNA editing enzyme-complex by hnRNP C1 protein and 40S hnRNP complexes. *Biol Chem* **379**, 1063-1073 (1998).
84. Yang, Y., Yang, Y. & Smith, H.C. Multiple protein domains determine the cell type-specific nuclear distribution of the catalytic subunit required for apolipoprotein B mRNA editing. *Proc Natl Acad Sci U S A* **94**, 13075-13080 (1997).
85. Navaratnam, N. et al. Escherichia coli cytidine deaminase provides a molecular model for ApoB RNA editing and a mechanism for RNA substrate recognition. *J Mol Biol* **275**, 695-714 (1998).
86. Teng, B.B. et al. Mutational analysis of apolipoprotein B mRNA editing enzyme (APOBEC1). structure-function relationships of RNA editing and dimerization. *J Lipid Res* **40**, 623-635 (1999).
87. Yang, Y., Sowden, M.P., Yang, Y. & Smith, H.C. Intracellular trafficking determinants in APOBEC-1, the catalytic subunit for cytidine to uridine editing of apolipoprotein B mRNA. *Exp Cell Res* **267**, 153-164 (2001).
88. Chester, A. et al. The apolipoprotein B mRNA editing complex performs a multifunctional cycle and suppresses nonsense-mediated decay. *EMBO J* **22**, 3971-3982 (2003).
89. Blanc, V., Kennedy, S. & Davidson, N.O. A novel nuclear localization signal in the auxiliary domain of apobec-1 complementation factor regulates nucleocytoplasmic import and shuttling. *J Biol Chem* **278**, 41198-41204 (2003).

90. Mukhopadhyay, D. et al. Thyroid hormone regulates hepatic triglyceride mobilization and apolipoprotein B messenger ribonucleic Acid editing in a murine model of congenital hypothyroidism. *Endocrinology* **144**, 711-719 (2003).
91. Lehmann, D.M., Galloway, C.A., Sowden, M.P. & Smith, H.C. Metabolic regulation of apoB mRNA editing is associated with phosphorylation of APOBEC-1 complementation factor. *Nucleic Acids Res* **34**, 3299-3308 (2006).
92. Lehmann, D.M. et al. Functional characterization of APOBEC-1 complementation factor phosphorylation sites. *Biochim Biophys Acta* **1773**, 408-418 (2007).
93. Giannoni, F. et al. Developmental regulation of the catalytic subunit of the apolipoprotein B mRNA editing enzyme (APOBEC-1) in human small intestine. *J Lipid Res* **36**, 1664-1675 (1995).
94. Greeve, J., Axelos, D., Welker, S., Schipper, M. & Greten, H. Distinct promoters induce APOBEC-1 expression in rat liver and intestine. *Arterioscler Thromb Vasc Biol* **18**, 1079-1092 (1998).
95. Hirano, K., Min, J., Funahashi, T. & Davidson, N.O. Cloning and characterization of the rat apobec-1 gene: a comparative analysis of gene structure and promoter usage in rat and mouse. *J Lipid Res* **38**, 1103-1119 (1997).
96. Qian, X., Balestra, M.E. & Innerarity, T.L. Two distinct TATA-less promoters direct tissue-specific expression of the rat apo-B editing catalytic polypeptide 1 gene. *J Biol Chem* **272**, 18060-18070 (1997).
97. Hirano, K., Min, J., Funahashi, T., Baunoch, D.A. & Davidson, N.O. Characterization of the human apobec-1 gene: expression in gastrointestinal tissues determined by alternative splicing with production of a novel truncated peptide. *J Lipid Res* **38**, 847-859 (1997).
98. Yamanaka, S. et al. Apolipoprotein B mRNA-editing protein induces hepatocellular carcinoma and dysplasia in transgenic animals. *Proc Natl Acad Sci U S A* **92**, 8483-8487 (1995).
99. Sowden, M., Hamm, J.K. & Smith, H.C. Overexpression of APOBEC-1 results in mooring sequence-dependent promiscuous RNA editing. *J Biol Chem* **271**, 3011-3017 (1996).
100. Yamanaka, S., Poksay, K.S., Driscoll, D.M. & Innerarity, T.L. Hyperediting of multiple cytidines of apolipoprotein B mRNA by APOBEC-1 requires auxiliary protein(s) but not a mooring sequence motif. *J Biol Chem* **271**, 11506-11510 (1996).
101. Yamanaka, S., Poksay, K.S., Arnold, K.S. & Innerarity, T.L. A novel translational repressor mRNA is edited extensively in livers containing tumors caused by the transgene expression of the apoB mRNA-editing enzyme. *Genes Dev* **11**, 321-333 (1997).
102. Harris, R.S., Petersen-Mahrt, S.K. & Neuberger, M.S. RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. *Mol Cell* **10**, 1247-1253 (2002).
103. Petersen-Mahrt, S.K. & Neuberger, M.S. In vitro deamination of cytosine to uracil in single-stranded DNA by apolipoprotein B editing complex catalytic subunit 1 (APOBEC1). *J Biol Chem* **278**, 19583-19586 (2003).
104. Dance, G.S. et al. Two proteins essential for apolipoprotein B mRNA editing are expressed from a single gene through alternative splicing. *J Biol Chem* **277**, 12703-12709 (2002).
105. Sowden, M.P., Lehmann, D.M., Lin, X., Smith, C.O. & Smith, H.C. Identification of novel alternative splice variants of APOBEC-1 complementation factor with different capacities to support apolipoprotein B mRNA editing. *J Biol Chem* **279**, 197-206 (2004).

106. Blanc, V. et al. Targeted deletion of the murine apobec-1 complementation factor (acf) gene results in embryonic lethality. *Mol Cell Biol* **25**, 7260-7269 (2005).
107. Snyder, E.M. et al. APOBEC1 complementation factor (A1CF) is dispensable for C-to-U RNA editing in vivo. *RNA* **23**, 457-465 (2017).
108. Fossat, N. et al. C to U RNA editing mediated by APOBEC1 requires RNA-binding protein RBM47. *EMBO Rep* **15**, 903-910 (2014).
109. Rosenberg, B.R., Hamilton, C.E., Mwangi, M.M., Dewell, S. & Papavasiliou, F.N. Transcriptome-wide sequencing reveals numerous APOBEC1 mRNA-editing targets in transcript 3' UTRs. *Nat Struct Mol Biol* **18**, 230-236 (2011).
110. Blanc, V. et al. Apobec1 complementation factor (A1CF) and RBM47 interact in tissue-specific regulation of C to U RNA editing in mouse intestine and liver. *RNA* **25**, 70-81 (2019).
111. Wolfe, A.D., Arnold, D.B. & Chen, X.S. Comparison of RNA Editing Activity of APOBEC1-A1CF and APOBEC1-RBM47 Complexes Reconstituted in HEK293T Cells. *J Mol Biol* **431**, 1506-1517 (2019).
112. Davies, M.S. et al. Sequence requirements for apolipoprotein B RNA editing in transfected rat hepatoma cells. *J Biol Chem* **264**, 13395-13398 (1989).
113. Driscoll, D.M., Wynne, J.K., Wallis, S.C. & Scott, J. An in vitro system for the editing of apolipoprotein B mRNA. *Cell* **58**, 519-525 (1989).
114. Tennyson, G.E., Sabatos, C.A., Higuchi, K., Meglin, N. & Brewer, H.B., Jr. Expression of apolipoprotein B mRNAs encoding higher- and lower-molecular weight isoproteins in rat liver and intestine. *Proc Natl Acad Sci U S A* **86**, 500-504 (1989).
115. Backus, J.W. & Smith, H.C. Three distinct RNA sequence elements are required for efficient apolipoprotein B (apoB) RNA editing in vitro. *Nucleic Acids Res* **20**, 6007-6014 (1992).
116. Richardson, N., Navaratnam, N. & Scott, J. Secondary structure for the apolipoprotein B mRNA editing site. Au-binding proteins interact with a stem loop. *J Biol Chem* **273**, 31707-31717 (1998).
117. Shah, R.R. et al. Sequence requirements for the editing of apolipoprotein B mRNA. *J Biol Chem* **266**, 16301-16304 (1991).
118. Maris, C., Masse, J., Chester, A., Navaratnam, N. & Allain, F.H. NMR structure of the apoB mRNA stem-loop and its interaction with the C to U editing APOBEC1 complementary factor. *RNA* **11**, 173-186 (2005).
119. Chen, S.H., Li, X.X., Liao, W.S., Wu, J.H. & Chan, L. RNA editing of apolipoprotein B mRNA. Sequence specificity determined by in vitro coupled transcription editing. *J Biol Chem* **265**, 6811-6816 (1990).
120. Backus, J.W. & Smith, H.C. Specific 3' sequences flanking a minimal apolipoprotein B (apoB) mRNA editing 'cassette' are critical for efficient editing in vitro. *Biochim Biophys Acta* **1217**, 65-73 (1994).
121. Driscoll, D.M., Lakhe-Reddy, S., Oleksa, L.M. & Martinez, D. Induction of RNA editing at heterologous sites by sequences in apolipoprotein B mRNA. *Mol Cell Biol* **13**, 7288-7294 (1993).
122. Higuchi, M. et al. RNA editing of AMPA receptor subunit GluR-B: a base-paired intron-exon structure determines position and efficiency. *Cell* **75**, 1361-1370 (1993).
123. Egebjerg, J., Kukekov, V. & Heinemann, S.F. Intron sequence directs RNA editing of the glutamate receptor subunit GluR2 coding sequence. *Proc Natl Acad Sci U S A* **91**, 10270-10274 (1994).

124. Kohler, M., Kornau, H.C. & Seeburg, P.H. The organization of the gene for the functionally dominant alpha-amino-3-hydroxy-5-methylisoxazole-4-propionic acid receptor subunit GluR-B. *J Biol Chem* **269**, 17367-17370 (1994).
125. Lomeli, H. et al. Control of kinetic properties of AMPA receptor channels by nuclear RNA editing. *Science* **266**, 1709-1713 (1994).
126. Yang, J.H., Sklar, P., Axel, R. & Maniatis, T. Editing of glutamate receptor subunit B pre-mRNA in vitro by site-specific deamination of adenosine. *Nature* **374**, 77-81 (1995).
127. Maas, S. et al. Structural requirements for RNA editing in glutamate receptor pre-mRNAs by recombinant double-stranded RNA adenosine deaminase. *J Biol Chem* **271**, 12221-12226 (1996).
128. Skuse, G.R., Cappione, A.J., Sowden, M., Metheny, L.J. & Smith, H.C. The neurofibromatosis type I messenger RNA undergoes base-modification RNA editing. *Nucleic Acids Res* **24**, 478-485 (1996).
129. Blanc, V. et al. Genome-wide identification and functional analysis of Apobec-1-mediated C-to-U RNA editing in mouse small intestine and liver. *Genome Biol* **15**, R79 (2014).
130. Soleymanjahi, S., Blanc, V. & Davidson, N. APOBEC1 mediated C-to-U RNA editing: target sequence and trans-acting factor contribution to 177 RNA editing events in 119 murine transcripts in-vivo. *RNA* **27**, 876-890 (2021).
131. Logue, E.C. et al. A DNA sequence recognition loop on APOBEC3A controls substrate specificity. *PLoS One* **9**, e97062 (2014).
132. Love, R.P., Xu, H. & Chelico, L. Biochemical analysis of hypermutation by the deoxycytidine deaminase APOBEC3A. *J Biol Chem* **287**, 30812-30822 (2012).
133. Herb, A., Higuchi, M., Sprengel, R. & Seeburg, P.H. Q/R site editing in kainate receptor GluR5 and GluR6 pre-mRNAs requires distant intronic sequences. *Proc Natl Acad Sci U S A* **93**, 1875-1880 (1996).
134. Burns, C.M. et al. Regulation of serotonin-2C receptor G-protein coupling by RNA editing. *Nature* **387**, 303-308 (1997).
135. Niswender, C.M., Sanders-Bush, E. & Emeson, R.B. Identification and characterization of RNA editing events within the 5-HT2C receptor. *Ann N Y Acad Sci* **861**, 38-48 (1998).
136. Polson, A.G. & Bass, B.L. Preferential selection of adenosines for modification by double-stranded RNA adenosine deaminase. *EMBO J* **13**, 5701-5711 (1994).
137. Lehmann, K.A. & Bass, B.L. Double-stranded RNA adenosine deaminases ADAR1 and ADAR2 have overlapping specificities. *Biochemistry* **39**, 12875-12884 (2000).
138. Eggington, J.M., Greene, T. & Bass, B.L. Predicting sites of ADAR editing in double-stranded RNA. *Nat Commun* **2**, 319 (2011).
139. Bazak, L. et al. A-to-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes. *Genome Res* **24**, 365-376 (2014).
140. Johnson, D.F., Poksay, K.S. & Innerarity, T.L. The mechanism for apo-B mRNA editing is deamination. *Biochem Biophys Res Commun* **195**, 1204-1210 (1993).
141. Macbeth, M.R. et al. Inositol hexakisphosphate is bound in the ADAR2 core and required for RNA editing. *Science* **309**, 1534-1539 (2005).
142. Matthews, M.M. et al. Structures of human ADAR2 bound to dsRNA reveal base-flipping mechanism and basis for site selectivity. *Nat Struct Mol Biol* **23**, 426-433 (2016).
143. Daujotyte, D. et al. HhaI DNA methyltransferase uses the protruding Gln237 for active flipping of its target cytosine. *Structure* **12**, 1047-1055 (2004).

144. Klimasauskas, S., Kumar, S., Roberts, R.J. & Cheng, X. HhaI methyltransferase flips its target base out of the DNA helix. *Cell* **76**, 357-369 (1994).
145. Kuttan, A. & Bass, B.L. Mechanistic insights into editing-site specificity of ADARs. *Proc Natl Acad Sci U S A* **109**, E3295-3304 (2012).
146. Thuy-Boun, A.S. et al. Asymmetric dimerization of adenosine deaminase acting on RNA facilitates substrate recognition. *Nucleic Acids Res* (2020).
147. Malik, T.N. et al. Regulation of RNA editing by intracellular acidification. *Nucleic Acids Res* **49**, 4020-4036 (2021).
148. Phelps, K.J. et al. Recognition of duplex RNA by the deaminase domain of the RNA editing enzyme ADAR2. *Nucleic Acids Res* **43**, 1123-1132 (2015).
149. Cho, D.S. et al. Requirement of dimerization for RNA editing activity of adenosine deaminases acting on RNA. *J Biol Chem* **278**, 17093-17102 (2003).
150. Poulsen, H. et al. Dimerization of ADAR2 is mediated by the double-stranded RNA binding domain. *RNA* **12**, 1350-1360 (2006).
151. Valente, L. & Nishikura, K. RNA binding-independent dimerization of adenosine deaminases acting on RNA and dominant negative effects of nonfunctional subunits on dimer functions. *J Biol Chem* **282**, 16054-16061 (2007).
152. Aydin, H., Taylor, M.W. & Lee, J.E. Structure-guided analysis of the human APOBEC3-HIV restrictome. *Structure* **22**, 668-684 (2014).
153. Salter, J.D. & Smith, H.C. Modeling the Embrace of a Mutator: APOBEC Selection of Nucleic Acid Ligands. *Trends Biochem Sci* **43**, 606-622 (2018).
154. Kouno, T. et al. Crystal structure of APOBEC3A bound to single-stranded DNA reveals structural basis for cytidine deamination and specificity. *Nat Commun* **8**, 15024 (2017).
155. Maiti, A. et al. Crystal Structure of a Soluble APOBEC3G Variant Suggests ssDNA to Bind in a Channel that Extends between the Two Domains. *J Mol Biol* **432**, 6042-6060 (2020).
156. Rathore, A. et al. The local dinucleotide preference of APOBEC3G can be altered from 5'-CC to 5'-TC by a single amino acid substitution. *J Mol Biol* **425**, 4442-4454 (2013).
157. Shi, K. et al. Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B. *Nat Struct Mol Biol* **24**, 131-139 (2017).
158. Ziegler, S.J. et al. Insights into DNA substrate selection by APOBEC3G from structural, biochemical, and functional studies. *PLoS One* **13**, e0195048 (2018).
159. Maiti, A. et al. Crystal structure of the catalytic domain of HIV-1 restriction factor APOBEC3G in complex with ssDNA. *Nat Commun* **9**, 2460 (2018).
160. Polevoda, B. et al. RNA binding to APOBEC3G induces the disassembly of functional deaminase complexes by displacing single-stranded DNA substrates. *Nucleic Acids Res* **43**, 9434-9445 (2015).
161. Prohaska, K.M., Bennett, R.P., Salter, J.D. & Smith, H.C. The multifaceted roles of RNA binding in APOBEC cytidine deaminase functions. *Wiley Interdiscip Rev RNA* **5**, 493-508 (2014).
162. Anant, S., MacGinnitie, A.J. & Davidson, N.O. apobec-1, the catalytic subunit of the mammalian apolipoprotein B mRNA editing enzyme, is a novel RNA-binding protein. *J Biol Chem* **270**, 14762-14767 (1995).
163. Li, J. et al. APOBEC3 multimerization correlates with HIV-1 packaging and restriction activity in living cells. *J Mol Biol* **426**, 1296-1307 (2014).
164. McDougall, W.M., Okany, C. & Smith, H.C. Deaminase activity on single-stranded DNA (ssDNA) occurs in vitro when APOBEC3G cytidine deaminase forms

homotetramers and higher-order complexes. *J Biol Chem* **286**, 30655-30661 (2011).
165. McDougall, W.M. & Smith, H.C. Direct evidence that RNA inhibits APOBEC3G ssDNA cytidine deaminase activity. *Biochem Biophys Res Commun* **412**, 612-617 (2011).
166. Wang, T. et al. 7SL RNA mediates virion packaging of the antiviral cytidine deaminase APOBEC3G. *J Virol* **81**, 13112-13124 (2007).
167. Bohn, M.F. et al. The ssDNA Mutator APOBEC3A Is Regulated by Cooperative Dimerization. *Structure* **23**, 903-911 (2015).
168. Woolf, T.M., Chase, J.M. & Stinchcomb, D.T. Toward the therapeutic editing of mutated RNA sequences. *Proc Natl Acad Sci U S A* **92**, 8298-8302 (1995).
169. Gronemeyer, T., Chidley, C., Juillerat, A., Heinis, C. & Johnsson, K. Directed evolution of O6-alkylguanine-DNA alkyltransferase for applications in protein labeling. *Protein Eng Des Sel* **19**, 309-316 (2006).
170. Juillerat, A. et al. Directed evolution of O6-alkylguanine-DNA alkyltransferase for efficient labeling of fusion proteins with small molecules in vivo. *Chem Biol* **10**, 313-317 (2003).
171. Juillerat, A. et al. Engineering substrate specificity of O6-alkylguanine-DNA alkyltransferase for specific protein labeling in living cells. *Chembiochem* **6**, 1263-1269 (2005).
172. Mollwitz, B. et al. Directed evolution of the suicide protein O(6)-alkylguanine-DNA alkyltransferase for increased reactivity results in an alkylated protein with exceptional stability. *Biochemistry* **51**, 986-994 (2012).
173. Goodtzova, K. et al. Repair of O6-benzylguanine by the Escherichia coli Ada and Ogt and the human O6-alkylguanine-DNA alkyltransferases. *J Biol Chem* **272**, 8332-8339 (1997).
174. Pegg, A.E. Repair of O(6)-alkylguanine by alkyltransferases. *Mutat Res* **462**, 83-100 (2000).
175. Gautier, A. et al. An engineered protein tag for multiprotein labeling in living cells. *Chem Biol* **15**, 128-136 (2008).
176. Los, G.V. et al. HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS Chem Biol* **3**, 373-382 (2008).
177. Damoiseaux, R., Keppler, A. & Johnsson, K. Synthesis and applications of chemical probes for human O6-alkylguanine-DNA alkyltransferase. *Chembiochem* **2**, 285-287 (2001).
178. Keppler, A. et al. A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nat Biotechnol* **21**, 86-89 (2003).
179. Gendreizig, S., Kindermann, M. & Johnsson, K. Induced protein dimerization in vivo through covalent labeling. *J Am Chem Soc* **125**, 14970-14971 (2003).
180. Stafforst, T. & Schneider, M.F. An RNA-deaminase conjugate selectively repairs point mutations. *Angew Chem Int Ed Engl* **51**, 11166-11169 (2012).
181. Vogel, P., Schneider, M.F., Wettengel, J. & Stafforst, T. Improving site-directed RNA editing in vitro and in cell culture by chemical modification of the guideRNA. *Angew Chem Int Ed Engl* **53**, 6267-6271 (2014).
182. Vogel, P. et al. Efficient and precise editing of endogenous transcripts with SNAP-tagged ADARs. *Nat Methods* **15**, 535-538 (2018).
183. Hanswillemenke, A., Kuzdere, T., Vogel, P., Jekely, G. & 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).

184. Vogel, P., Hanswillemenke, A. & Stafforst, T. Switching Protein Localization by Site-Directed RNA Editing under Control of Light. *ACS Synth Biol* **6**, 1642-1649 (2017).
185. Stroppel, A.S., Lappalainen, R. & Stafforst, T. Controlling Site-Directed RNA Editing by Chemically Induced Dimerization. *Chemistry* **27**, 12300-12304 (2021).
186. Antony, J.S. et al. Gene correction of HBB mutations in CD34(+) hematopoietic stem cells using Cas9 mRNA and ssODN donors. *Mol Cell Pediatr* **5**, 9 (2018).
187. Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A. & Liu, D.R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420-424 (2016).
188. Porto, E.M., Komor, A.C., Slaymaker, I.M. & Yeo, G.W. Base editing: advances and therapeutic opportunities. *Nat Rev Drug Discov* **19**, 839-859 (2020).
189. Abudayyeh, O.O. et al. C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **353**, aaf5573 (2016).
190. Cox, D.B.T. et al. RNA editing with CRISPR-Cas13. *Science* **358**, 1019-1027 (2017).
191. Abudayyeh, O.O. et al. A cytosine deaminase for programmable single-base RNA editing. *Science* **365**, 382-386 (2019).
192. Azad, M.T.A., Bhakta, S. & Tsukahara, T. Site-directed RNA editing by adenosine deaminase acting on RNA for correction of the genetic code in gene therapy. *Gene Ther* **24**, 779-786 (2017).
193. Bhakta, S., Azad, M.T.A. & Tsukahara, T. Genetic code restoration by artificial RNA editing of Ochre stop codon with ADAR1 deaminase. *Protein Eng Des Sel* **31**, 471-478 (2018).
194. Bhakta, S., Sakari, M. & Tsukahara, T. RNA editing of BFP, a point mutant of GFP, using artificial APOBEC1 deaminase to restore the genetic code. *Sci Rep* **10**, 17304 (2020).
195. Huang, X. et al. Programmable C-to-U RNA editing using the human APOBEC3A deaminase. *EMBO J* **39**, e104741 (2020).
196. Katrekar, D. et al. In vivo RNA editing of point mutations via RNA-guided adenosine deaminases. *Nat Methods* **16**, 239-242 (2019).
197. Katrekar, D. et al. Comprehensive interrogation of the ADAR2 deaminase domain for engineering enhanced RNA editing activity and specificity. *Elife* **11** (2022).
198. Katrekar, D. et al. Efficient in vitro and in vivo RNA editing via recruitment of endogenous ADARs using circular guide RNAs. *Nat Biotechnol* **40**, 938-945 (2022).
199. Li, J., Fan, G., Sakari, M. & Tsukahara, T. Improvement of C-to-U RNA editing using an artificial MS2-APOBEC system. *Biotechnol J*, e2300321 (2023).
200. Litke, J.L. & Jaffrey, S.R. Highly efficient expression of circular RNA aptamers in cells using autocatalytic transcripts. *Nat Biotechnol* **37**, 667-675 (2019).
201. Merkle, T. et al. Precise RNA editing by recruiting endogenous ADARs with antisense oligonucleotides. *Nat Biotechnol* **37**, 133-138 (2019).
202. Montiel-Gonzalez, M.F., Vallecillo-Viejo, I., Yudowski, G.A. & Rosenthal, J.J. Correction of mutations within the cystic fibrosis transmembrane conductance regulator by site-directed RNA editing. *Proc Natl Acad Sci U S A* **110**, 18285-18290 (2013).
203. Montiel-Gonzalez, M.F., Vallecillo-Viejo, I.C. & Rosenthal, J.J. An efficient system for selectively altering genetic information within mRNAs. *Nucleic Acids Res* **44**, e157 (2016).

204. Qu, L. et al. Programmable RNA editing by recruiting endogenous ADAR using engineered RNAs. *Nat Biotechnol* **37**, 1059-1069 (2019).
205. Rauch, S. et al. Programmable RNA-Guided RNA Effector Proteins Built from Human Parts. *Cell* **178**, 122-134 e112 (2019).
206. Reautschnig, P. et al. CLUSTER guide RNAs enable precise and efficient RNA editing with endogenous ADAR enzymes in vivo. *Nat Biotechnol* (2022).
207. Sinnamon, J.R. et al. Site-directed RNA repair of endogenous Mecp2 RNA in neurons. *Proc Natl Acad Sci U S A* **114**, E9395-E9402 (2017).
208. Tohama, T., Sakari, M. & Tsukahara, T. Development of a Single Construct System for Site-Directed RNA Editing Using MS2-ADAR. *Int J Mol Sci* **21** (2020).
209. Wettengel, J., Reautschnig, P., Geisler, S., Kahle, P.J. & Stafforst, T. Harnessing human ADAR2 for RNA repair - Recoding a PINK1 mutation rescues mitophagy. *Nucleic Acids Res* **45**, 2797-2808 (2017).
210. Kannan, S. et al. Compact RNA editors with small Cas13 proteins. *Nat Biotechnol* (2021).
211. Konermann, S. et al. Transcriptome Engineering with RNA-Targeting Type VI-D CRISPR Effectors. *Cell* **173**, 665-676 e614 (2018).
212. Liu, Y. et al. REPAIRx, a specific yet highly efficient programmable A > I RNA base editor. *EMBO J* **39**, e104748 (2020).
213. Li, G. et al. Developing PspCas13b-based enhanced RESCUE system, eRESCUE, with efficient RNA base editing. *Cell Commun Signal* **19**, 84 (2021).
214. Austin, R.J., Xia, T., Ren, J., Takahashi, T.T. & Roberts, R.W. Designed arginine-rich RNA-binding peptides with picomolar affinity. *J Am Chem Soc* **124**, 10966-10967 (2002).
215. Baron-Benhamou, J., Gehring, N.H., Kulozik, A.E. & Hentze, M.W. Using the lambdaN peptide to tether proteins to RNAs. *Methods Mol Biol* **257**, 135-154 (2004).
216. Chattopadhyay, S., Garcia-Mena, J., DeVito, J., Wolska, K. & Das, A. Bipartite function of a small RNA hairpin in transcription antitermination in bacteriophage lambda. *Proc Natl Acad Sci U S A* **92**, 4061-4065 (1995).
217. Bardwell, V.J. & Wickens, M. Purification of RNA and RNA-protein complexes by an R17 coat protein affinity method. *Nucleic Acids Res* **18**, 6587-6594 (1990).
218. Charlesworth, C.T. et al. Identification of preexisting adaptive immunity to Cas9 proteins in humans. *Nat Med* **25**, 249-254 (2019).
219. Han, W. et al. Programmable RNA base editing with a single gRNA-free enzyme. *Nucleic Acids Res* **50**, 9580-9595 (2022).
220. Wang, M., Oge, L., Perez-Garcia, M.D., Hamama, L. & Sakr, S. The PUF Protein Family: Overview on PUF RNA Targets, Biological Functions, and Post Transcriptional Regulation. *Int J Mol Sci* **19** (2018).
221. Stein, C.A. et al. Efficient gene silencing by delivery of locked nucleic acid antisense oligonucleotides, unassisted by transfection reagents. *Nucleic Acids Res* **38**, e3 (2010).
222. Deleavey, G.F. & Damha, M.J. Designing chemically modified oligonucleotides for targeted gene silencing. *Chem Biol* **19**, 937-954 (2012).
223. Khvorova, A. & Watts, J.K. The chemical evolution of oligonucleotide therapies of clinical utility. *Nat Biotechnol* **35**, 238-248 (2017).
224. Roberts, T.C., Langer, R. & Wood, M.J.A. Advances in oligonucleotide drug delivery. *Nat Rev Drug Discov* **19**, 673-694 (2020).
225. Crooke, S.T., Wang, S., Vickers, T.A., Shen, W. & Liang, X.H. Cellular uptake and trafficking of antisense oligonucleotides. *Nat Biotechnol* **35**, 230-237 (2017).

226. Wan, W.B. & Seth, P.P. The Medicinal Chemistry of Therapeutic Oligonucleotides. *J Med Chem* **59**, 9645-9667 (2016).
227. Rohner, E., Yang, R., Foo, K.S., Goedel, A. & Chien, K.R. Unlocking the promise of mRNA therapeutics. *Nat Biotechnol* **40**, 1586-1600 (2022).
228. Haque, A. et al. Chemically modified hCFTR mRNAs recuperate lung function in a mouse model of cystic fibrosis. *Sci Rep* **8**, 16776 (2018).
229. Doherty, E.E. et al. Rational Design of RNA Editing Guide Strands: Cytidine Analogs at the Orphan Position. *J Am Chem Soc* **143**, 6865-6876 (2021).
230. Inoue, H., Hayase, Y., Iwai, S. & Ohtsuka, E. Sequence-dependent hydrolysis of RNA using modified oligonucleotide splints and RNase H. *FEBS Lett* **215**, 327-330 (1987).
231. Guschlbauer, W. & Jankowski, K. Nucleoside conformation is determined by the electronegativity of the sugar substituent. *Nucleic Acids Res* **8**, 1421-1433 (1980).
232. Inoue, H. et al. Synthesis and hybridization studies on two complementary nona(2'-O-methyl)ribonucleotides. *Nucleic Acids Res* **15**, 6131-6148 (1987).
233. Freier, S.M. & Altmann, K.H. The ups and downs of nucleic acid duplex stability: structure-stability studies on chemically-modified DNA:RNA duplexes. *Nucleic Acids Res* **25**, 4429-4443 (1997).
234. Hayase, Y., Inoue, H. & Ohtsuka, E. Secondary structure in formylmethionine tRNA influences the site-directed cleavage of ribonuclease H using chimeric 2'-O-methyl oligodeoxyribonucleotides. *Biochemistry* **29**, 8793-8797 (1990).
235. Geary, R.S. et al. Pharmacokinetic properties of 2'-O-(2-methoxyethyl)-modified oligonucleotide analogs in rats. *J Pharmacol Exp Ther* **296**, 890-897 (2001).
236. Scharner, J. & Aznarez, I. Clinical Applications of Single-Stranded Oligonucleotides: Current Landscape of Approved and In-Development Therapeutics. *Mol Ther* **29**, 540-554 (2021).
237. Kawasaki, A.M. et al. Uniformly modified 2'-deoxy-2'-fluoro phosphorothioate oligonucleotides as nuclease-resistant antisense compounds with high affinity and specificity for RNA targets. *J Med Chem* **36**, 831-841 (1993).
238. Manoharan, M. 2'-carbohydrate modifications in antisense oligonucleotide therapy: importance of conformation, configuration and conjugation. *Biochim Biophys Acta* **1489**, 117-130 (1999).
239. Allerson, C.R. et al. Fully 2'-modified oligonucleotide duplexes with improved in vitro potency and stability compared to unmodified small interfering RNA. *J Med Chem* **48**, 901-904 (2005).
240. Springer, A.D. & Dowdy, S.F. GalNAc-siRNA Conjugates: Leading the Way for Delivery of RNAi Therapeutics. *Nucleic Acid Ther* **28**, 109-118 (2018).
241. Braasch, D.A. & Corey, D.R. Locked nucleic acid (LNA): fine-tuning the recognition of DNA and RNA. *Chem Biol* **8**, 1-7 (2001).
242. Koshkin, A. et al. Synthesis of the adenine, cytosine, guanine, 5-methylcytosine, thymine and uracil bicyclonucleoside monomers, oligomerisation, and unprecedented nucleic acid recognition. *Tetrahedron* **54**, 3607-3630 (1998).
243. Obika, S. et al. Synthesis of 2′-O, 4′-C-methyleneuridine and-cytidine. Novel bicyclic nucleosides having a fixed C3,-endo sugar puckering. *Tetrahedron Letters* **38**, 8735-8738 (1997).
244. Gaus, H.J. et al. Characterization of the interactions of chemically-modified therapeutic nucleic acids with plasma proteins using a fluorescence polarization assay. *Nucleic Acids Res* **47**, 1110-1122 (2019).

245. Miller, C.M. et al. Stabilin-1 and Stabilin-2 are specific receptors for the cellular internalization of phosphorothioate-modified antisense oligonucleotides (ASOs) in the liver. *Nucleic Acids Res* **44**, 2782-2794 (2016).
246. Eckstein, F. Developments in RNA chemistry, a personal view. *Biochimie* **84**, 841-848 (2002).
247. Iwamoto, N. et al. Control of phosphorothioate stereochemistry substantially increases the efficacy of antisense oligonucleotides. *Nat Biotechnol* **35**, 845-851 (2017).
248. Byrne, M. et al. Stereochemistry Enhances Potency, Efficacy, and Durability of Malat1 Antisense Oligonucleotides In Vitro and In Vivo in Multiple Species. *Transl Vis Sci Technol* **10**, 23 (2021).
249. Alterman, J.F. et al. Hydrophobically Modified siRNAs Silence Huntingtin mRNA in Primary Neurons and Mouse Brain. *Mol Ther Nucleic Acids* **4**, e266 (2015).
250. Braasch, D.A. et al. RNA interference in mammalian cells by chemically-modified RNA. *Biochemistry* **42**, 7967-7975 (2003).
251. Hassler, M.R. et al. Comparison of partially and fully chemically-modified siRNA in conjugate-mediated delivery in vivo. *Nucleic Acids Res* **46**, 2185-2196 (2018).
252. Haraszti, R.A. et al. 5´-Vinylphosphonate improves tissue accumulation and efficacy of conjugated siRNAs in vivo. *Nucleic Acids Res* **45**, 7581-7592 (2017).
253. Parmar, R. et al. 5'-(E)-Vinylphosphonate: A Stable Phosphate Mimic Can Improve the RNAi Activity of siRNA-GalNAc Conjugates. *Chembiochem* **17**, 985-989 (2016).
254. Prakash, T.P. et al. Identification of metabolically stable 5'-phosphate analogs that support single-stranded siRNA activity. *Nucleic Acids Res* **43**, 2993-3011 (2015).
255. Prakash, T.P. et al. Targeted delivery of antisense oligonucleotides to hepatocytes using triantennary N-acetyl galactosamine improves potency 10-fold in mice. *Nucleic Acids Res* **42**, 8796-8807 (2014).
256. Tanowitz, M. et al. Asialoglycoprotein receptor 1 mediates productive uptake of N-acetylgalactosamine-conjugated and unconjugated phosphorothioate antisense oligonucleotides into liver hepatocytes. *Nucleic Acids Res* **45**, 12388-12400 (2017).
257. Nair, J.K. et al. Multivalent N-acetylgalactosamine-conjugated siRNA localizes in hepatocytes and elicits robust RNAi-mediated gene silencing. *J Am Chem Soc* **136**, 16958-16961 (2014).
258. Nishina, K. et al. Efficient In Vivo Delivery of siRNA to the Liver by Conjugation of alpha-Tocopherol. *Mol Ther* **16**, 734-740 (2008).
259. Krieg, A.M. et al. Modification of antisense phosphodiester oligodeoxynucleotides by a 5' cholesteryl moiety increases cellular association and improves efficacy. *Proc Natl Acad Sci U S A* **90**, 1048-1052 (1993).
260. Wolfrum, C. et al. Mechanisms and optimization of in vivo delivery of lipophilic siRNAs. *Nat Biotechnol* **25**, 1149-1157 (2007).
261. Soutschek, J. et al. Therapeutic silencing of an endogenous gene by systemic administration of modified siRNAs. *Nature* **432**, 173-178 (2004).
262. Eder, P.S., DeVine, R.J., Dagle, J.M. & Walder, J.A. Substrate specificity and kinetics of degradation of antisense oligonucleotides by a 3' exonuclease in plasma. *Antisense Res Dev* **1**, 141-151 (1991).
263. Watt, A.T., Swayze, G., Swayze, E.E. & Freier, S.M. Likelihood of Nonspecific Activity of Gapmer Antisense Oligonucleotides Is Associated with Relative Hybridization Free Energy. *Nucleic Acid Ther* **30**, 215-228 (2020).

264. Ostergaard, M.E. et al. Rational design of antisense oligonucleotides targeting single nucleotide polymorphisms for potent and allele selective suppression of mutant Huntingtin in the CNS. *Nucleic Acids Res* **41**, 9634-9650 (2013).
265. Koller, E. et al. Mechanisms of single-stranded phosphorothioate modified antisense oligonucleotide accumulation in hepatocytes. *Nucleic Acids Res* **39**, 4795-4807 (2011).
266. Yu, C., Brussaard, A.B., Yang, X., Listerud, M. & Role, L.W. Uptake of antisense oligonucleotides and functional block of acetylcholine receptor subunit gene expression in primary embryonic neurons. *Dev Genet* **14**, 296-304 (1993).
267. Hug, N., Longman, D. & Caceres, J.F. Mechanism and regulation of the nonsense-mediated decay pathway. *Nucleic Acids Res* **44**, 1483-1495 (2016).
268. Vogel, P. & Stafforst, T. Site-directed RNA editing with antagomir deaminases--a tool to study protein and RNA function. *ChemMedChem* **9**, 2021-2025 (2014).
269. Liu, C.C., Liu, C.C., Kanekiyo, T., Xu, H. & Bu, G. Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. *Nat Rev Neurol* **9**, 106-118 (2013).
270. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**, 3406-3415 (2003).
271. Morin, P.J. et al. Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC. *Science* **275**, 1787-1790 (1997).
272. Zhan, T., Rindtorff, N. & Boutros, M. Wnt signaling in cancer. *Oncogene* **36**, 1461-1473 (2017).
273. Ayanoglu, F.B., Elcin, A.E. & Elcin, Y.M. Bioethical issues in genome editing by CRISPR-Cas9 technology. *Turk J Biol* **44**, 110-120 (2020).
274. Berk, C. et al. Pharmacodynamic and Pharmacokinetic Properties of Full Phosphorothioate Small Interfering RNAs for Gene Silencing In Vivo. *Nucleic Acid Ther* (2020).
275. Nair, J.K. et al. Impact of enhanced metabolic stability on pharmacokinetics and pharmacodynamics of GalNAc-siRNA conjugates. *Nucleic Acids Res* **45**, 10969-10977 (2017).

5. Appendix

5.1. Materials and Methods

(snap)/(snap)$_2$-guideRNA synthesis

guideRNAs were purchased from Eurogentec or BioSpring. guideRNAs either contained a C6-Amino linker at their 5'-end or a C7-Amino linker at their 3'-end. 120 nmol of snap-linker were preactivated for 60 min. at 30°C with 224 nmol 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDCI, Alfa Aesat, #A10807), 306 nmol of N-hydroxysuccinimide (NHS, Sigma-Aldrich, #130672), and 2 µl of a 5% solution of N,N-diisopropylethylamine (DIPEA, abcr GmbH, #AB182190) in dimethyl sulfoxide (DMSO, Carl Roth, #A994.1). For conjugation, half of the preactivation mix was added to 8.5 µl of NH2-guideRNA (50 µg of 6 µg/µl in H_2O) containing 4 µl of DIPEA solution (5% in DMSO) and incubated at 30°C. After 30 min., the rest of the preactivation mix was added and incubated for 60 min. at 30°C.

For bivalent guideRNAs, 120 nmol of (snap)2-linker was preactivated for 4 h at 45°C or over-night at 37°C with 540 nmol N,N'-diisopropylcarbodiimide (DIC, Sigma-Aldrich, #D125407-5g), 920 nmol of NHS, and 2 µl of a 5% DIPEA solution in DMSO. Residual DIC was removed by lyophilization. Product was dissolved in 12 µl of a 1.7 % DIPEA solution in DMSO, added to 8.5 µl of NH_2-guideRNA solution (50 µg of 6 µg/µl in H_2O), and incubated at 37°C for 2 h.

Linker-bearing guideRNAs were purified by urea polyacrylamide gel electrophoresis (PAGE). 0.8 mm thick 20% 5M urea polyacrylamide gels were cast in large glass plates. For this, 204 ml of ROTHIPHORESE®Sequencing gel concentrate (25 %, Carl Roth, #3043.1), 13 ml of ROTHIPHORESE®Sequencing gel buffer concentrate (50 %, Carl Roth, #3050.1), 13 ml of H_2O, 650 µl of Ammonium peroxydisulphate (Carl Roth, #9592.3, 10% solution in H_20), and 65 µl of tetramethylehtylenediamine (TEMED; Carl Roth, #2367.3) were mixed and subsequently poured into the glass plates. After solidification over-night at room temperature gels were ready to run. Samples were mixed with 5 µl of loading dye containing bromphenol blue sodium salt (Carl Roth, #A512.1) and xylene cyanole (Carl Roth, #A513.1) dissolved in a 1:10 dilution of ROTHIPHORESE® Sequencing gel diluent (50 %, Carl Roth, #3047.1) in H2O. PAGE was conducted in Tris-Boric Acid-EDTA buffer (TBE, Tris 8.9 mM, Boric Acid 8.9 mM, EDTA 0.2 mM) for 6 h, at 65 W, 90 mA. Correct bands were excised, and gel slices were incubated in nuclease-free H2O over-night at 4°C

while vigorously shaking. For purification 0.1 volumes of sodium acetate (3 M in H_2O) was added to the guideRNAs together with 3.5 volumes of ethanol. After over-night precipitation at -20°C, samples were centrifuged for 60 min at 4°C at 14,000 RPM. Supernatant was aspirated and guideRNA pellet was washed with 500 µl cold ethanol (70% in H_2O). After centrifugation for 60 min at 4°C with 14,000 RPM, the pellet was dissolved in nuclease-free H_2O. The concentration was measured by NanoDrop™ 2000/2000c Spectrophotometers (ThermoFisher Scientific).

Cy5 labeling of guideRNAs

For some stability assays, NH-guideRNAs were labeled with Cy5 for detection without the use of staining with SYBRGold. To this end, 1 nmol of NH-guideRNAs were coupled over-night at room-temperature in darkness with Cy5-NHS ester (0.5 mM final concentration) in reaction buffer consisting of a 1:20 dilution of $NaHCO_3$ (0.2 M, pH = 9) with PBS. guideRNAs were purified by ethanol precipitation (see prior section).

Creation of cell lines stably expressing SNAP-ADARs

Cell lines were created by utilizing the Flp-In™ T-REx™ system (Invitrogen, #R78007). For cultivation HEK 293 Flp-In™ T-REx™ cells were held in Dulbecco's modified eagle medium (DMEM; Life Technologies, #41965062) containing fetal bovine serum (FBS; Life Technologies, #10270, 10% final conc.), Zeocin™ Selection Reagent (Z; Invitrogen, #R25001, 100 µg/ml final conc.), and Blasticidin S Hydrochloride (B; Blasticidin S Hydrochloride, Carl Roth, #CP14.2, 15 µg/ml final conc.) Cell lines were generated as already described.[182] 1.6 x 10^6 cells were incubated in a 6 cm dish in DMEM +10% FBS, +Z, +B for 24 h. Before transfection, medium was changed to antibiotic-free DMEM +10% FBS and cells were incubated for 1h. Cells were then transfected with 3.6 µg of pOG44 Flp-Recombinase expression vector (invitrogen, #V600520) and 0.4 µg of pcDNA™5/FRT Mammalian Expression Vector (Invitrogen, #V601020) with 12 µl Lipofectamine™ 2000 (ThermoFisher Scientific, #11668019) all dissolved in 600µl Opti-minimal essential medium I (Opti-MEM™ I; Life Technologies, #11058021). 24 h post-infection, correctly generated cells were selected for by replacing Zeocin by CELLPURE® Hygromycin B solution (H; Carl Roth, #CP12.1, 100 µg/ml final conc.) Two-weeks after selection start, cells were taken into permanent culture in DMEM +10% FBS, +B, +H.

Creation of cell lines stably expressing mMeCP2-eGFP constructs

Cells stably and constitutively expressing mMeCP2-eGFP constructs were created with the PiggyBac transposon system from HEK 293 Flp-In™ T-REx™ cells, stably expressing SNAP-ADAR variants under doxycycline induction. For this, 1 x 10^5 HEK 293 Flp-In™ T-REx™ cells were incubated in a 24-well plate for 24 h in DMEM + 10% FBS, +B, +H. Before transfection, medium was changed to DMEM +10% FBS. Cells were transfected with 0.75 µg of plasmid carrying one of the mMeCP2-eGFP variants and a puromycin resistance gene, 0.25 µg of plasmid for transposase expression, and 3 µl of FuGENE® 6 (Promega, #E2691) all dissolved in Opti-MEM™. 24 h after transfection medium was exchanged by DMEM +10% FBS, puromycin dihydrochloride (2 µg/ml final conc., Carl Roth, #0240.1). After 48 h, cells were washed with PBS and medium was replaced with DMEM +10% FBS, +B, +H.

Editing experiments

For full induction of transgene expression, HEK 293 Flp-In™ T-Rex™ cells stably expressing SNAP-ADARs were incubated for 24 hours with 10 ng/ml doxycycline (AppliChem, #A2951,005). 8 x 10^4 cells were transfected with 5 pmol of snap-guideRNA or 1 pmol of $(snap)_2$-guideRNA using 0.75 µl Lipofectamine™ 2000 by dropwise pipetting cell suspension onto the guideRNA/Lipofectamine mix in a 96-well format. Unless differently noted, editing endpoint was 24 h after transfection.

For full induction of transgene expression in HeLa cells stably expressing SNAP-ADAR1Q were incubated for 24 hours with 1 µg/µl doxycycline. 5 x 10^4 cells were transfected with 2 pmol of $(snap)_2$-guideRNA using 0.75 µl Lipofectamine™ 2000 by dropwise pipetting cell suspension onto the guideRNA/Lipofectamine mix in a 96-well format. Every 24 hours, cells were split and editing yield was assessed. Endpoint was 72 hours post initial guideRNA transfection.

Editing yield was assessed as follows. At endpoint, RLT-buffer (Qiagen, #79216) was used to lyse the cells. Total RNA was then isolated with the Monarch® RNA Cleanup Kit 10 µg (NewEngland BioLabs, #T2030L) following manufacturer's instructions. DNA was then depleted by incubation of 1 µg of total RNA with DNase I (NewEngland BioLabs, #M0303L) at 37°C for 30 minutes. DNase I was inactivated in presence of EDTA (2 µl, 25 mM final conc.) by incubation at 75°C for 10 minutes. 250 ng (7.5 µl) of total RNA was reverse transcribed and the target site was amplified with the One*Taq*® One-Step RT-PCR

Kit (NewEngland BioLabs, #E5315S) in 25 µl total volume. PCR products were excised and isolated using the NucleoSpin® Gel and PCR Clean-up Mini kit (Macherey-Nagel, #740609) following manufacturer's instructions. 100 ng of PCR product was sequenced by Sanger Sequencing carried out by Microsynth or Eurofins.

Gymnotic uptake experiments in HeLa cells

Gymnotic uptake experiments with HeLa cells stably expressing SNAP-ADAR1Q under doxycycline-induction was conducted in different ways. For the experiments with constant medium or medium exchange, 1.25 x 10^4 HeLa cells were seeded in a 96-well format and incubated with doxycycline (1 µg/µl final conc.) for 24 h to induce SNAP-ADAR1Q expression. 24 h after seeding guideRNAs were dissolved in DMEM +10% FBS, + doxycycline (2 µg/µl final conc.) and added to the cells at indicated final concentrations. For the samples with the medium exchange, fresh DMEM +FBS (10 % final conc.), +doxycycline (2 µg/µl final conc.) containing guideRNAs at indicated final concentrations to the cells every 24 hours. 72 h post initial guideRNA supplementation, total RNA was isolated and editing yield was assessed, as described above.

For the gymnotic uptake experiments over seven days, 2 x 104 HeLa cells stably expressing SNAP-ADAR1Q under doxycycline-induction were seeded in a 48-well format and incubated for 24 h in DMEM +FBS (10 % final conc.), +doxycycline (2 µg/µl final conc.) to induce SNAP-ADAR1Q expression. 24 h after seeding, guideRNA dissolved in DMEM, +FBS (10 % final conc.), +doxycycline (2 µg/µl final conc.) at indicated final concentrations were added to the cells. At indicated time-points cells reached confluence and were thus split and reseeded. 24 hours after seeding, fresh DMEM +FBS (10 % final conc.), +doxycycline (2 µg/µl final conc.) containing guideRNAs at indicated final concentrations were added to the cells. When split, a part of the cells was lysed and the total RNA was isolated, as described above. Editing yield was then assessed. as described above.

Stability in Serum and Tritosomes

FBS was diluted in phosphate-buffer saline (PBS; 10% final concentration). 180 pmol of NH-/Cy5-guideRNA were incubated in PBS containing FBS at a final concentration of 9% FBS at 37°C. At indicated timepoints 10 µl (= 15 pmol guideRNA) were sampled and snap-frozen in liquid nitrogen and subsequently stored at -80°C. For stability in pure serum, 180 pmol of either NH-guideRNA or $(snap)_2$-guideRNA were incubated in FBS at a final

concentration of 90% FBS at 37°C. At indicated timepoints 10 µl (= 15 pmol guideRNA) were sampled and snap-frozen in liquid nitrogen and subsequently stored at -80°C.

For stability in rat liver lysosomal extract, tritosome (Xenotech LLC, #R0610.LT) stock solution was diluted to 0.5 U/ml acid phosphatase final concentration using 20 mM citrate buffer (pH = 5), as described elsewhere.[253, 274, 275] 12 µl of a 15 pmol/µl dilution of either NH-guideRNA or $(snap)_2$-guideRNA were incubated in 108 µl of diluted tritosomes at 37°C. At indicated timepoints 10 µl (= 15 pmol guideRNA) were sampled and snap-frozen in liquid nitrogen and subsequently stored at -80°C.

Analysis was conducted by 20% 5M urea PAGE, as described above. Samples incubated in 90% FBS or tritosomes, were treated with Proteinase K at 50°C for 10 min. Afterwards, samples were supplemented with loading dye and loaded on the gel. The gel was run in TBE-buffer for 6 h at 65 W and 90 mA. At the end of the run, the gel was stained by with SYBRGold dissolved in TBE-buffer for 20 min. at room temperature. When guideRNAs were Cy5-labeled, fluorescent bands were detected at 646 nm wavelength.

5.2. guideRNA Sequences

Internal number	NUMBER	NAME	TRANSCRIPT	Context 5'	CHANGE	COMPANY	EPSILON	LENGTH / nt	5' Mod	SEQUENCE	3'Mod
85	1	GAPDH_ORF_2_PTO	GAPDH	UAG	L249L	BioSpring	227(+2.5)	22	snap	mU*mA*mU mGmGmU mUmUmU mU rC **rC** rA mGmAmC mGmG*mC* mA*mG*mG	
146	2	GAPDH_BG85_2OMe	GAPDH	UAG	L249L	BioSpring	227(+2.5)	22	snap	mU*mA*mU mGmGmU mUmUmU mU mC **rC** mA mGmAmC mGmG*mC* mA*mG*mG	
147	3	GAPDH_BG85_F	GAPDH	UAG	L249L	BioSpring	227(+2.5)	22	snap	mU*mA*mU mGmGmU mUmUmU mU fC **rC** fA mGmAmC mGmG*mC* mA*mG*mG	
148	4	GAPDH_BG85_5P_Ome	GAPDH	UAG	L249L	BioSpring	227(+2.5)	22	snap	mU*mA*mU mGmGmU mUmUmU mU mC* **rC** mA mGmAmC mGmG*mC* mA*mG*mG	
149	5	GAPDH_BG85_5P_F	GAPDH	UAG	L249L	BioSpring	227(+2.5)	22	snap	mU*mA*mU mGmGmU mUmUmU mU fC* **rC** fA mGmAmC mGmG*mC* mA*mG*mG	
150	6	GAPDH_BG85_3P_Ome	GAPDH	UAG	L249L	BioSpring	227(+2.5)	22	snap	mU*mA*mU mGmGmU mUmUmU mU mC **rC*** mA mGmAmC mGmG*mC* mA*mG*mG	
151	7	GAPDH_BG85_3P_F	GAPDH	UAG	L249L	BioSpring	227(+2.5)	22	snap	mU*mA*mU mGmGmU mUmUmU mU fC **rC*** fA mGmAmC mGmG*mC* mA*mG*mG	
152	8	GAPDH_BG85_3xP	GAPDH	UAG	L249L	BioSpring	227(+2.5)	22	snap	mU*mA*mU mGmGmU mUmUmU mU rC* **rC*** rA* mGmAmC mGmG*mC* mA*mG*mG	
164	9	dC_Ome_BG85_GAPDH	GAPDH	UAG	L249L	Eurogentec	219(+2.5)	22	snap	mU*mA*mU mGmGmU mUmUmU mU mC **dC** mA mGmAmC mGmG*mC* mA*mG*mG	
161	10	dC_BG85_GAPDH	GAPDH	UAG	L249L	Eurogentec	219(+2.5)	22	snap	mU*mA*mU mGmGmU mUmUmU mU rC **dC** rA mGmAmC mGmG*mC* mA*mG*mG	
162	11	2dC_BG85_GAPDH	GAPDH	UAG	L249L	Eurogentec	219(+2.5)	22	snap	mU*mA*mU mGmGmU mUmUmU mU dC **dC** rA mGmAmC mGmG*mC* mA*mG*mG	
163	12	dA_2dC_BG85_GAPDH	GAPDH	UAG	L249L	Eurogentec	219(+2.5)	22	snap	mU*mA*mU mGmGmU mUmUmU mU dC **dC** dA mGmAmC mGmG*mC* mA*mG*mG	
193	13	GAPDH_BG85_OMeCs_dCs_OMeAs	GAPDH	UAG	L249L	BioSpring	227(+2.5)	22	snap	mU*mA*mU mGmGmU mUmUmU mU mC* **rC*** mA* mGmAmC mGmG*mC* mA*mG*mG	
194	14	GAPDH_BG85_2xFs_Cs	GAPDH	UAG	L249L	BioSpring	227(+2.5)	22	snap	mU*mA*mU mGmGmU mUmUmU mU fC* **dC*** fA* mGmAmC mGmG*mC* mA*mG*mG	
192	15	GAPDH_BG85_dCs_dCs_dAs	GAPDH	UAG	L249L	BioSpring	227(+2.5)	22	snap	mU*mA*mU mGmGmU mUmUmU mU dC* **dC*** dA* mGmAmC mGmG*mC* mA*mG*mG	
106	16	GAPDH_ORF_UAU	GAPDH	UAU	I30V	BioSpring	224(+2.5)	22	snap	mC*mU*mA mGmGmC mAmAmC mA rA **rC** rA mUmCmC mAmC*mU* mU*mU*mA	
169	17	GAPDH_BG-106_3P_OMe	GAPDH	UAU	I30V	Eurogentec	218.5(+2.5)	22	snap	mC*mU*mA mGmGmC mAmAmC mA mA **rC*** mA mUmCmC mAmC*mU* mU*mU*mA	
199	18	GAPDH_BG-106_dA_dC	GAPDH	UAU	I30V	BioSpring	224(+2.5)	22	snap	mC*mU*mA mGmGmC mAmAmC mA dA **dC** rA mUmCmC mAmC*mU* mU*mU*mA	
200	19	GAPDH_BG-106_dA_dC_OMeA	GAPDH	UAU	I30V	BioSpring	224(+2.5)	22	snap	mC*mU*mA mGmGmC mAmAmC mA dA **dC** mA mUmCmC mAmC*mU* mU*mU*mA	
105	20	GAPDH_ORF_UAC	GAPDH	UAC	T211A	BioSpring	222(+2.5)	22	snap	mC*mC*mG mAmGmC mGmCmC mA rG **rC** rA mGmAmG mGmC*mA* mG*mG*mG	
168	21	GAPDH_BG-105_3P_OMe	GAPDH	UAC	T211A	Eurogentec	213(+2.5)	22	snap	mC*mC*mG mAmGmC mGmCmC mA mG **rC*** mA mGmAmG mGmC*mA* mG*mG*mG	

198	22	GAPDH_BG-105_dG_dC_OMeA	GAPDH	UAC	T211A	BioSpring	222(+2.5)	22	snap	mC*mC*mG mAmGmC mGmCmC mA dG **dC** mA mGmAmG mGmC*mA* mG*mG*mG	
195	23	GAPDH_BG-105_dG_dC	GAPDH	UAC	T211A	BioSpring	222(+2.5)	22	snap	mC*mC*mG mAmGmC mGmCmC mA dG **dC** rA mGmAmG mGmC*mA* mG*mG*mG	
196	24	GAPDH_BG-105_dG_dCs	GAPDH	UAC	T211A	BioSpring	222(+2.5)	22	snap	mC*mC*mG mAmGmC mGmCmC mA dG **dC*** rA mGmAmG mGmC*mA* mG*mG*mG	
197	25	GAPDH_BG-105_dG_dC_As	GAPDH	UAC	T211A	BioSpring	222(+2.5)	22	snap	mC*mC*mG mAmGmC mGmCmC mA dG **dC** rA* mGmAmG mGmC*mA* mG*mG*mG	
101	26	GAPDH_ORF_AAG	GAPDH	AAG	K27K	BioSpring	226(+2.5)	22	snap	mU*mG*mU mAmUmA mUmCmC mA rC **rC** rU mUmAmC mCmA*mG* mA*mG*mU	
166	27	GAPDH_BG-101_3P_OMe	GAPDH	AAG	K27K	Eurogentec	219(+2.5)	22	snap	mU*mG*mU mAmUmA mUmCmC mA mC **rC*** mU mUmAmC mCmA*mG* mA*mG*mU	
224	28	GAPDH_BG-101_dC_cs_dU	GAPDH	AAG	K27K	BioSpring	226(+2.5)	22	snap	mU*mG*mU mAmUmA mUmCmC mA dC **rC*** dU mUmAmC mCmA*mG* mA*mG*mU	
225	29	GAPDH_BG-101_fC_cs_fU	GAPDH	AAG	K27K	BioSpring	226(+2.5)	22	snap	mU*mG*mU mAmUmA mUmCmC mA fC **rC*** fU mUmAmC mCmA*mG* mA*mG*mU	
205	30	GAPDH_BG-101_2xOMes_Cs	GAPDH	AAG	K27K	BioSpring	226(+2.5)	22	snap	mU*mG*mU mAmUmA mUmCmC mA mC* **rC*** mU* mUmAmC mCmA*mG* mA*mG*mU	
117	31	GAPDH_ORF_CAC	GAPDH	CAC	A158A	BioSpring	224(+2.5)	22	snap	mA*mA*mC mGmCmC mAmGmG mG rG **rC** rG mCmUmA mAmG*mC* mA*mG*mU	
172	32	GAPDH_BG-117_3P_OMe	GAPDH	CAC	A158A	Eurogentec	217(+2.5)	22	snap	mA*mA*mC mGmCmC mAmGmG mG mG **rC*** mG mCmUmA mAmG*mC* mA*mG*mU	
285	33	NL_mMeCP2_W104X	mMeCP2-GFP	UAG	X104W	Eurogentec	234(+5)	25	(snap)2	mU*mC*mG lTmUlT mCmGmU mGmU rC **rC** rA mAmCmC mUmUmC mA*mG*lG* mC*lA	
289	34	NL_mMeCP2_W104X_OMeC_rCs_OMeA	mMeCP2-GFP	UAG	X104W	Eurogentec	209(+5)	22	(snap)2	lT*mU*lT mCmGmU mGmU mC **dC*** mA mAmCmC mUmUmC mA*mG*lG* mC*lA	
287	35	NL_mMeCP2_W104X_2dC_rA	mMeCP2-GFP	UAG	X104W	Eurogentec	209(+5)	22	(snap)2	lT*mU*lT mCmGmU mGmU dC **dC** rA mAmCmC mUmUmC mA*mG*lG* mC*lA	
288	36	NL_mMeCP2_W104X_2dC_dA	mMeCP2-GFP	UAG	X104W	Eurogentec	209(+5)	22	(snap)2	lT*mU*lT mCmGmU mGmU dC **dC** dA mAmCmC mUmUmC mA*mG*lG* mC*lA	
290	37	NL_mMeCP2_W104X_2dC_rA-Chol	mMeCP2-GFP	UAG	X104W	Eurogentec	236(+5)	25	(snap)2	mU*mC*mG lTmUlT mCmGmU mGmU dC **dC** rA mAmCmC mUmUmC mA*mG*lG* mC*lA	Cholesteryl-TEG
286	38	NL_mMeCP2_R106Q	mMeCP2-GFP	CAA	Q106R	Eurogentec	243(+5)	25	(snap)2	mA*mC*mA lTmUlA mAmGmC mUmU rU **rC** rG mUmGmU mCmCmA mA*mC*lC* mU*lT	
294	39	NL_mMeCP2_R106Q_OMeU_rC_rG	mMeCP2-GFP	CAA	Q106R	Eurogentec	243(+5)	25	(snap)2	mA*mC*mA lTmUlA mAmGmC mUmU mU **rC** rG mUmGmU mCmCmA mA*mC*lC* mU*lT	
292	40	NL_mMeCP2_R106Q_dT_dC_dG	mMeCP2-GFP	CAA	Q106R	Eurogentec	243(+5)	25	(snap)2	mA*mC*mA lTmUlA mAmGmC mUmU dT **dC** dG mUmGmU mCmCmA mA*mC*lC* mU*lT	
293	41	NL_mMeCP2_R106Q_dT_dC_rG	mMeCP2-GFP	CAA	Q106R	Eurogentec	243(+5)	25	(snap)2	mA*mC*mA lTmUlA mAmGmC mUmU dT **dC** rG mUmGmU mCmCmA mA*mC*lC* mU*lT	
295	42	NL_mMeCP2_R106Q_rU_rC_rI	mMeCP2-GFP	CAA	Q106R	Eurogentec	239(+5)	25	(snap)2	mA*mC*mA lTmUlA mAmGmC mUmU rU **rC** rI mUmGmU mCmCmA mA*mC*lC* mU*lT	

311	43	NL_mMeCP2_W104X_2dC_dA_25nt	mMeCP2-GFP	UAG	X104W	Eurogentec	236(+5)	25	(snap)2	mU*mC*mG lTmUlT mCmGmU mGmU dC **dC** dA mAmCmC mUmUmC mA*mG*lG* mC*lA	
312	44	NL_mMeCP2_W104X_2dC_dA_25nt_Chol	mMeCP2-GFP	UAG	X104W	Eurogentec	236(+5)	25	(snap)2	mU*mC*mG lTmUlT mCmGmU mGmU dC **dC** dA mAmCmC mUmUmC mA*mG*lG* mC*lA	Cholesteryl-TEG
314	45	NL_mMeCP2_W104X_2dC_dA_all-PS	mMeCP2-GFP	UAG	X104W	Eurogentec	209(+5)	22	(snap)2	lT*mU*lT* mC*mG*mU* mG*mU* dC* **dC*** dA* mA*mC*mC* mU*mU*mC* mA*mG*lG* mC*lA	
313	46	NL_mMeCP2_W104X_2dC_dA_all-PS_C	mMeCP2-GFP	UAG	X104W	Eurogentec	209(+5)	22	(snap)2	lT*mU*lT* mC*mG*mU* mG*mU* dC* **dC*** dA* mA*mC*mC* mU*mU*mC* mA*mG*lG* mC*lA	Cholesteryl-TEG
307	47	NL_mMeCP2_W104X_2dC_dA_3'PS	mMeCP2-GFP	UAG	X104W	Eurogentec	209(+5)	22	(snap)2	lT*mU*lT mCmGmU mGmU dC **dC** dA mA*mC*mC* mU*mU*mC* mA*mG*lG* mC*lA	
308	48	NL_mMeCP2_W104X_2dC_dA_3'PS_C	mMeCP2-GFP	UAG	X104W	Eurogentec	209(+5)	22	(snap)2	lT*mU*lT mCmGmU mGmU dC **dC** dA mA*mC*mC* mU*mU*mC* mA*mG*lG* mC*lA	Cholesteryl-TEG
309	49	NL_mMeCP2_W104X_2dC_dA_5'+3'PS	mMeCP2-GFP	UAG	X104W	Eurogentec	209(+5)	22	(snap)2	lT*mU*lT* mC*mG*mU* mG*mU* dC **dC** dA mA*mC*mC* mU*mU*mC* mA*mG*lG* mC*lA	
310	50	NL_mMeCP2_W104X_2dC_dA_5'+3'PS_C	mMeCP2-GFP	UAG	X104W	Eurogentec	209(+5)	22	(snap)2	lT*mU*lT* mC*mG*mU* mG*mU* dC **dC** dA mA*mC*mC* mU*mU*mC* mA*mG*lG* mC*lA	Cholesteryl-TEG
315	51	NL_mMeCP2_W104X_2dC_dA_8-1-11	mMeCP2-GFP	UAG	X104W	Eurogentec	186(+5)	20	(snap)2	lT*mU*lC mGmUmG mU dC **dC** dA mAmCmC mUmUmC* mA*lG*mG* lC	
316	52	NL_mMeCP2_W104X_2dC_dA_7-1-10	mMeCP2-GFP	UAG	X104W	Eurogentec	170(+5)	18	(snap)2	lT*mC*lG mUmGmU dC **dC** dA mAmCmC mUmU*mC* lA*mG*lG	
317	53	NL_mMeCP2_W104X_2dC_dA_6-1-9	mMeCP2-GFP	UAG	X104W	Eurogentec	150(+5)	16	(snap)2	lC*mG*lT mGmU dC **dC** dA mAmCmC mU*mU*lC* mA*lG	
318	54	NL_mMeCP2_W104X_2dC_dA_9-1-9	mMeCP2-GFP	UAG	X104W	Eurogentec	179(+5)	19	(snap)2	lT*mU*lT mCmGmU mGmU dC **dC** dA mAmCmC mU*mU*lC* mA*lG	
319	55	NL_mMeCP2_W104X_2dC_dA_8-1-8	mMeCP2-GFP	UAG	X104W	Eurogentec	160(+5)	17	(snap)2	lT*mU*lC mGmUmG mU dC **dC** dA mAmCmC* mU*lT*mC* lA	
320	56	NL_mMeCP2_W104X_2dC_dA_7-1-7	mMeCP2-GFP	UAG	X104W	Eurogentec	137(+5)	15	(snap)2	lT*mC*lG mUmGmU dC **dC** dA mAmC*mC* lT*mU*lC	
346	57	NL_mMeCP2_W104X_2dC_dA_6-1-9 all PS 5'NH	mMeCP2-GFP	UAG	X104W	Eurogentec	150(+5)	16	(snap)2	lC*mG*lT* mG*mU* dC* **dC*** dA* mA*mC*mC* mU*mU*lC* mA*lG	
347	58	NL_mMeCP2_W104X_2dC_dA_9-1-6 all PS 3'NH	mMeCP2-GFP	UAG	X104W	Eurogentec	149(+5)	16		lT*mU*lT* mC*mG*mU* mG*mU* dC* **dC*** dA* mA*mC* lC*mU*lT	(snap)2
348	59	GAPDH_ORF#2_6-1-9_all PS_2dC_dA	GAPDH	UAG	L249L	Eurogentec	157(+5)	16	(snap)2	lT*mU*lT* mU*mU* dC* **dC*** dA* mG*mA*mC* mG*mG*lC* mA*lG	
180	60	Stat1_Y701C_rA_rC_rA	STAT1	UAU	Y701C	Eurogentec		25	(snap)2	mA*mG*mU lGmUlC mUmUmG mAmU rA **rC** rA mUmCmC mAmGmU mU*mC*lC* mU*lT	
508	61	Stat1_Y701C_dA_dC_dA_all PS_6-1-9_5end PS	STAT1	UAU	Y701C	Eurogentec	158(+5)	16	(snap)2	*lT*mU*lG* mA*mU* dA* **dC*** dA* mU*mC*mC* mA*mG*lT* mU*lC	

509	62	Stat1_Y701C_dA_dC_dA_ all PS_6-1-9_5end dT	STAT1	UAU	Y701C	Eurogentec	166(+5)	17	(snap)2	amino-dT *lT*mU*lG* mA*mU* dA* **dC*** dA* mU*mC*mC* mA*mG*lT* mU*lC	
510	63	Stat1_Y701C_dA_dC_dA_ all PS_9-1-6_3End PS	STAT1	UAU	Y701C	Eurogentec	164(+5)	16		lG*mU*lC* mU*mU*mG* mA*mU* dA* **dC*** dA* mU*mC*lC* mA*lG*	(snap)2
511	64	Stat1_Y701C_dA_dC_dA_ all PS_9-1-6_3End dT	STAT1	UAU	Y701C	Eurogentec	173(+5)	17		lG*mU*lC* mU*mU*mG* mA*mU* dA* **dC*** dA* mU*mC*lC* mA*lG* dT-amino	(snap)2

5.3. Other contributions

5.3.1. Patent applications

Eberhard Karls Universität Tübingen (2023). **Antisense Oligonucleotides (ASO) for Efficient and Precise RNA Editing with Endogenous Adenosine Deaminase Acting on RNA (ADAR)**. Inventors: Thorsten Stafforst, **Ngadhnjim Latifi**, Laura Sophia Pfeiffer. 8.6.2023. Application: 29.11.2022. PCT, disclosure document (Offenlegungsschrift) WO 2023/099494 A1

Eberhard Karls Universität Tübingen. **Chemically modified antisense oligonucleotides (ASOS) and compositions comprising the same for RNA editing**. Inventors: Thorsten Stafforst, Laura Sophia Pfeiffer, **Ngadhnjim Latifi**. Application: 30.11.2022. PCT, application document (Anmeldeschrift) PCT/EP2022/083943

5.3.2. Talks

Simultaneous Site-directed A-to-I and C-to-U Editing (short poster talk)
Ngadhnjim Latifi, Anna S. Stroppel, Aline M. Mack, Thorsten Stafforst, *10th German nucleic acids chemistry meeting*, 15-17 September 2021, *Bad Herrenalb, Germany,*

Programmable Tools for Transcript Engineering
Ngadhnjim Latifi, Thorsten Stafforst, *SPP-1784 Meeting*, 30 August – 01 September 2021, *Mainz*

5.3.3. Poster Presentations

Targeted A-to-I and C-to-U Editing using SNAP-fused effectors
Ngadhnjim Latifi, Aline M. Mack, Clemens Lochmann, Irem Tellioglu, F. Nina Papavasiliou, Thorsten Stafforst, *73rd Mosbach colloquium*, 30 March-02 April 2022, *Mosbach, Germany,*

Simultaneous Site-directed A-to-I and C-to-U Editing
Ngadhnjim Latifi*, Anna S. Stroppel*, Aline M. Mack, Thorsten Stafforst, *10th German nucleic acids chemistry meeting*, 15-17 September 2021, *Bad Herrenalb, Germany, *contributed equally*

Optimizing the gRNA chemistry for Site-directed RNA Editing

Ngadhnjim Latifi, Clemens Lochmann, Aline M. Mack, Tobias Merkle, Mareike Volz, Karthika D. Selvasaravanan, Alfred Hanswillemenke, Madeleine Heep, Thorsten Stafforst, *15th Annual Meeting of the Oligonucleotide Therapeutics Society*, 13-16 October 2019, *München, Germany*

Optimizing the gRNA chemistry for the SNAP-ADAR system

Ngadhnjim Latifi*, Paul Vogel, Madeleine Heep, Karthika D. Selvasaravanan, Alfred Hanswillemenke, Tobias Merkle, Thorsten Stafforst, 4th *PhD/PostDoc-Retreat, Interfaculty Institute of Biochemistry*, 09-10 May 2019, *Freudenstadt, Germany*

***Also involved as co-organizer of the retreat.**

Orthogonal protein tags – Expanding the RNA editing platform

Anna S. Stroppel, **Ngadhnjim Latifi**, Alfred Hanswillemenke, Cindy Odenwald, Aline M. Mack, Thorsten Stafforst, 2nd kick-off Meeting of the SPP1784, 10-12 December 2018, *Mainz, Germany*

Optimizing the gRNA chemistry for the SNAP-ADAR system

Ngadhnjim Latifi, Paul Vogel, Madeleine Heep, Karthika D Selvasaravanan, Alfred Hanswillemenke, Tobias Merkle, Thorsten Stafforst, 3rd *PhD/PostDoc-Retreat, Interfaculty Institute of Biochemistry*, 26-27 July 2018, *Hechingen, Germany*

5.4. Manuscripts

5.4.1. Manuscript 1 (published)

Precise and efficient C-to-U RNA base editing with SNAP-CDAR-S

Ngadhnjim Latifi, Aline Maria Mack, Irem Tellioglu, Salvatore Di Giorgio and Thorsten Stafforst, *Nucleic Acids Research*, 2023, *51.15,* e84

Published online 18 July 2023

Nucleic Acids Research, 2023, Vol. 51, No. 15 ***e84***
https://doi.org/10.1093/nar/gkad598

Precise and efficient C-to-U RNA base editing with SNAP-CDAR-S

Ngadhnjim Latifi [1], Aline Maria Mack [1], Irem Tellioglu [2,3], Salvatore Di Giorgio [2] and Thorsten Stafforst [1,4,*]

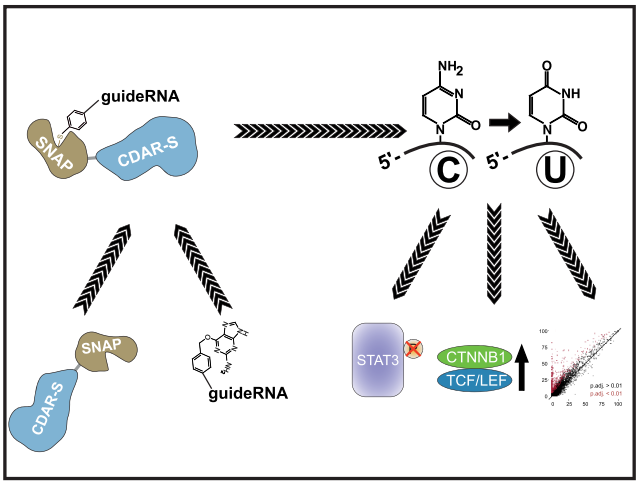
[1]Interfaculty Institute of Biochemistry, University of Tübingen, Auf der Morgenstelle 15, 72076 Tübingen, Germany, [2]Division of Immune Diversity (D150), German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany, [3]Faculty of Engineering, University of Heidelberg, 69120 Heidelberg, Germany and [4]Gene and RNA Therapy Center (GRTC), Faculty of Medicine University Tuebingen, Germany

Received February 20, 2023; Revised May 08, 2023; Editorial Decision June 27, 2023; Accepted July 05, 2023

ABSTRACT

Site-directed RNA base editing enables the transient and dosable change of genetic information and represents a recent strategy to manipulate cellular processes, paving ways to novel therapeutic modalities. While tools to introduce adenosine-to-inosine changes have been explored quite intensively, the engineering of precise and programmable tools for cytidine-to-uridine editing is somewhat lacking behind. Here we demonstrate that the cytidine deaminase domain evolved from the ADAR2 adenosine deaminase, taken from the RESCUE-S tool, provides very efficient and highly programmable editing when changing the RNA targeting mechanism from Cas13-based to SNAP-tag-based. Optimization of the guide RNA chemistry further allowed to dramatically improve editing yields in the difficult-to-edit 5′-CCN sequence context thus improving the substrate scope of the tool. Regarding editing efficiency, SNAP-CDAR-S outcompeted the RESCUE-S tool clearly on all tested targets, and was highly superior in perturbing the β-catenin pathway. NGS analysis showed similar, moderate global off-target A-to-I and C-to-U editing for both tools.

GRAPHICAL ABSTRACT



INTRODUCTION

Cytidine (C) deamination yielding uridine (U) is a well-known posttranscriptional reaction that diversifies genetic information at the RNA level (1). The enzymatic base conversion is carried out by hydrolases/deaminases belonging to the class of AID/APOBEC proteins, of which some are specific for RNA, while others can use both RNA and DNA, or only DNA as substrates. The first C-to-U RNA editing enzyme described was APOBEC1 (APO1) (2), which catalyzes the switch from the long ApoB100 to the short ApoB48 isoform by rewriting a glutamine codon (5′-CAA) into a STOP codon (5′-UAA) in the enterocytes of the small intestine (3). Later, single-strand RNA editing activity of further members of the APOBEC family, including APOBEC3A and 3G, was discovered. However, the biological function and targets of their RNA editing activity remained unclear to some extent (1). C-to-U RNA editing activity is typically found only in a sub-set of tissues, like small intestine and liver for APO1, or monocytes and

*To whom correspondence should be addressed. Tel: +49 707129 75376; Email: thorsten.stafforst@uni-tuebingen.de

macrophages for A3A, but can be up- and down-regulated in various pathologic situations and play a role in tumor evolution (4), for example (5). AID/APOBEC enzymes are often recruited to their targets by the help of auxiliary proteins, e.g. RBM47 (6) and A1CF (7) for APO1, and have a very strong and thus limiting preference for specific dinucleotides as editing substrates (1). Highly edited substrates, like the Glutamine-to-STOP site in the ApoB transcript, are placed in specific secondary structures that assist the recruitment and activity of the deaminase (8,9).

Targeted RNA base editing aims at harnessing C-to-U and A(denosine)-to-I(nosine) editing activity for the rewriting of genetic information, including the substitution of amino acids and the formation (C-to-U) or removal (A-to-I) of premature STOP codons (10). The approach opens novel avenues for drug discovery, promising to bypass technical and ethical issues related to genome editing (10). In this field, our lab contributed an RNA-targeting platform based on fusion proteins of the self-labeling SNAP-tag (11) (Figure 1A) (12). Initially, we engineered a programmable A-to-I RNA base editor by fusing the SNAP-tag (11) with the catalytic domain of the RNA editing enzyme ADAR (adenosine deaminase acting on RNA) (13,14). In these fusions, the SNAP-tag exploits its self-labeling activity to covalently tether to a guideRNA in a defined 1:1 stoichiometry by recognizing a benzylguanine (BG) moiety, the so-called self-labeling moiety, at the guideRNA. According to simple Watson-Crick base-pairing rules, the guideRNA addresses the editing of one specific adenosine residue in a selected transcript with high efficiency, broad codon scope, and very good precision (14). Competing RNA-targeting platforms have been developed based on Cas proteins (15), or trans-tethering approaches (16). While each approach has its specific strength and weakness (10,12), a clear advantage of the SNAP-tag approach is its human origin, its small size, the ease of transfecting of the chemically stabilized guideRNA(s) (14), the possibility of concurrent editing (14), and the ready inclusion of small molecule (17) and photo control (18,19). Furthermore, we have recently shown the concurrent and fully orthogonal usage of two independent RNA editing effectors by complementing a SNAP-tagged C-to-U editing effector with a HALO-tagged A-to-I editing tool within the same cell (20). In the latter study, we exploited the C-to-U deaminase domain from murine APO1. Other labs have developed C-to-U RNA base editing effectors based on human APO1 or APO3A. In the first example, RNA-targeting was based on the trans-tethering approach with the MS2/MCP system (21). In the latter, the dCas13 platform was applied (22). However, none of these approaches is yet working optimally. Our approach with SNAP-tagged APO1 (20) gave low editing yields on endogenous targets and its programmability, which is the addressing of any given target cytidine with a guide RNA, was somehow limited due to the strong requirement (8) for APO1 substrates to be located in specific secondary structures. While this was better solved for the A3A target (22), this tool suffers from the strong substrate codon preference for 5′-UC. An exciting alternative came from the engineering of an artificial C-to-U editing enzyme. Specifically, laboratory evolution was used to engineer the A-to-I deaminase domain of the hyperactive E488Q mutant (23) of the ancestor ADAR2 into a C-to-U editing enzyme (24). With a dCas13-based RNA-targeting mechanism the so-called RESCUE tool was steered to its target RNAs. While the programmability was good, the editing yields remained moderate for 5′-WC (W = A or U) codons and low for 5′-CC, whereas 5′-GC codons were hardly editable (24), mirroring the well-known codon preference (25) of ADAR2. Furthermore, the RESCUE tool retained notable A-to-I off-target editing beside C-to-U off-target editing. A high-fidelity variant, RESCUE-S, was developed (23), that carried an additional point mutation. However, the point mutation lowered both, the C-to-U on-target and the A-to-I off-target editing yields.

Here, we now show that the high-fidelity cytidine deaminase acting on RNA (CDAR) domain from the RESCUE-S tool works very well when we replace the dCas13 domain with a SNAP-tag for RNA targeting. In particular on endogenous transcripts, the SNAP-CDAR-S outcompetes the RESCUE-S tool clearly and achieves moderate to good editing yields for all 5′-HC codons (H = C, A, U) under very good control of A-to-I and C-to-U bystander editing.

MATERIALS AND METHODS

Generation of guideRNAs

guideRNAs (gRNA) were designed and purchased from Eurogentec or Sigma-Aldrich with a 5′-C6-Amino linker or a 3′-C7-Amino linker referred to as NH_2-gRNA. snap/$(snap)_2$-gRNA synthesis was carried out as previously described (20). Briefly, snap-linker was pre-activated with EDCI for 60 min at 30°C and $(snap)_2$-linker was pre-activated with DIC for 4 h at 45°C or over-night at 37°C. Coupling of snap-linker to gRNA was carried out at 30°C for 90 min and coupling of $(snap)_2$-Linker was carried out at 37°C for 2 h. Purification of gRNA was carried out by 5M urea PAGE and subsequent ethanol precipitation. Concentration and purity were determined by NanoDrop™ 2000/2000c Spectrophotometers (ThermoFisher Scientific). A detailed protocol is also given in the Supplementary Information (pages 3–4).

Generation of stable cell lines and editing

All transgenic cell lines were created from parental HEK 293 Flp-In™ T-REx™ cells (Invitrogen). Parental cells were cultivated in Dulbecco's modified Eagle's medium (DMEM; Life Technologies) with fetal bovine serum (FBS; Life Technologies, 10% final conc.), 100 μg/ml Zeocin (Z; ThermoFisher) and 15 μg/ml Blasticidin (B; Blasticidin S Hydrochloride, Carl Roth) 37°C and 5% CO_2. For generation of transgenic cell lines, 1.6×10^6 of parental cells were transfected with 4 μg of a 9:1 ratio of pOG44 Flp-Recombinase expression vector (Invitrogen) and pcDNA™5/FRT Mammalian Expression Vector (Invitrogen) bearing the transgene of interest using Lipofectamine™ 2000 (Life Technologies). For successful generation of transgenic cells, cells were selected for two weeks with DMEM, FBS (10%), B, 100 μg/ml Hygromycin (H; Carl Roth). Cells were then kept in DMEM, FBS (10%), B, H at 37°C and 5% CO_2. For SNAP-editase expressing cells, 3×10^5 cells were seeded

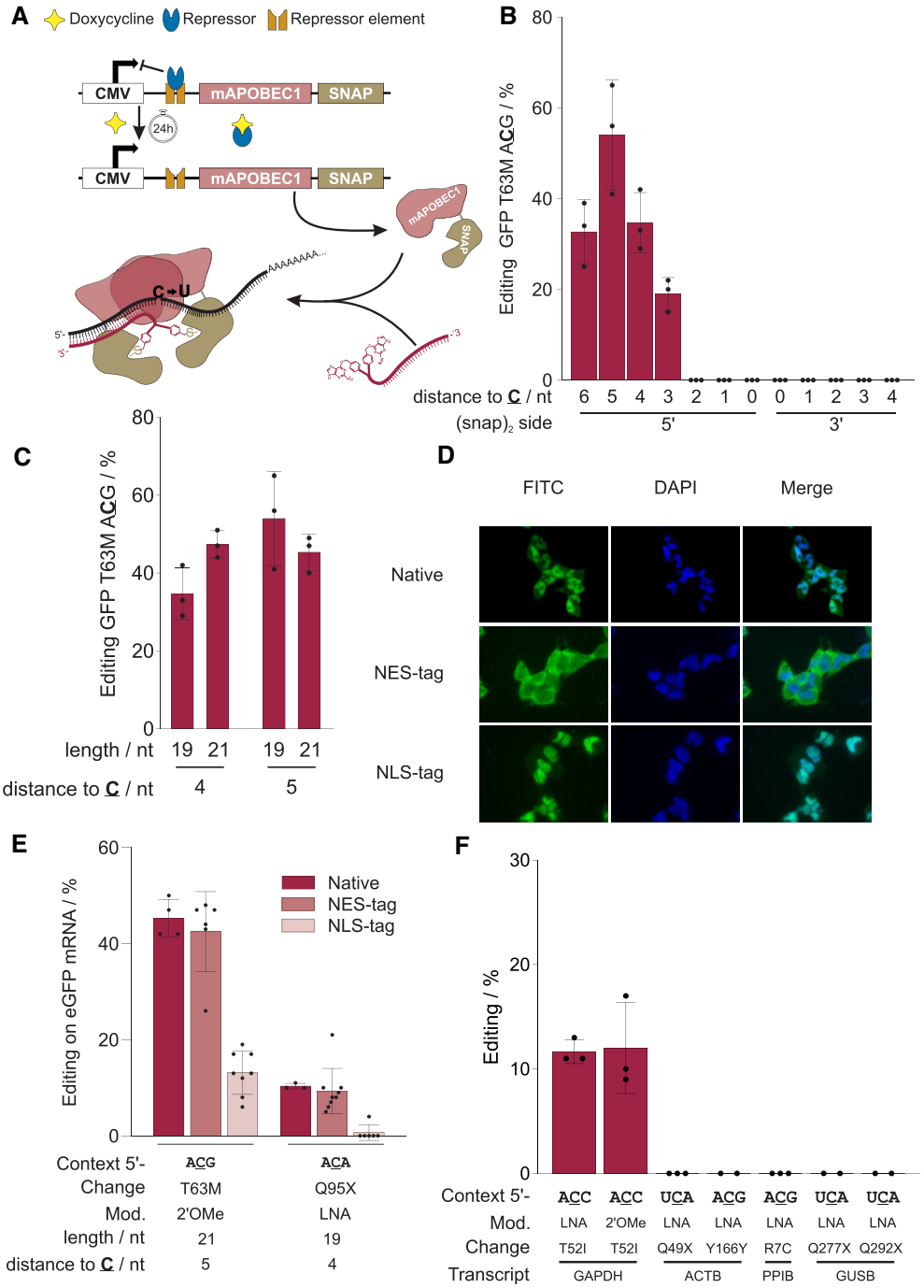


Figure 1. Properties of the APO1S tool. (**A**) Scheme of the doxycycline-inducible APO1S tool and the guide RNA-dependent editing reaction. (**B**) Dependency of editing yield of the T63M site (5′-ACG) in the eGFP reporter gene on the positioning of the guide RNA (19 nucleotides length, fully 2′OMe modified) relative to the target site ('C'). (**C**) Fine-tuning of guide RNA length (19, 21 nt) and positioning (4 or 5 nt 5′ to the target cytidine) for optimal editing performance. (**D**) Analysis of APO1S transgene expression and localization by fluorescence microscopy. The native mApobec1 sequence was either amended with an NES or NLS tag, as indicated, leading to cytosolic or nuclear localization of the APO1S protein, which was stained with SNAP-tag-reactive BG-FITC (green channel). (**E**) Effect of editase localization on editing of two sites in an eGFP reporter with the respective best performing guideRNA design. (**F**) Editing performance of guide RNAs (21 nt, position 5) with different backbone chemistries (2′-OMe, LNA) inducing the indicated amino acid changes at the respective endogenous transcripts demonstrating limited programmability of the tool. Data in (B), (C), (E) and (F) are shown as the mean ± s.d. of $N \geq 2$ independent experiments as represented by individual data points.

in DMEM, FBS (10%) and doxycycline (Dox; PanReacAppliChem, 10 ng/ml final conc.) in a 24-well format for transgene induction. 24 h after seeding cells were transfected with 300 ng of pcDNA3.1 expressing eGFP or pDNA expressing APOE4 (Addgene, #87087) with 0.9 μl of Lipofectamine™ 2000 (Life Technologies) per 300 ng of pDNA. 24h post-transfection, 8×10^4 cells were transfected with 5 pmol (unless differently stated) of snap/$(snap)_2$-gRNA with 0.75 μl of Lipofectamine™ 2000 per 5 pmol of gRNA. Unless differently stated, endpoint was 48 h post gRNA transfection. If endogenous transcripts were targeted, pDNA expression was omitted. For RESCUE-S expressing cells, 2×10^4 cells were seeded in a 96-well format in DMEM, FBS (10%) and doxycycline (10 ng/ml final conc.) for transgene induction. 24 h later, cells were transfected similarly to as previously described. Briefly, cells were transfected with 300 ng of gRNA expressing pDNA and if applicable 40 ng of reporter pDNA with 0.5 μl of Lipofectamine™ 2000. Unless differently stated, endpoint was 48 h post transfection.

Editing yield analysis

At endpoint, cells were lysed in 50 μl per well (96-well format) RLT buffer (Qiagen). Total RNA was isolated using the Monarch® RNA Cleanup Kit 10 μg (New England BioLabs) following manufacturer's instructions. Target sites were amplified using either One Step RT-PCR Kit (BiotechRabbit) or One*Taq*® One-Step RT-PCR Kit (New England BioLabs) and the appropriate primers. Sanger sequencing was performed by Microsynth or Eurofins. Editing yield was determined as the ratio of peak heights at target sites in the chromatogram of samples.

Microscopy

5×10^4 Flp-In™ T-Rex™ cells expressing SNAP-editases were seeded on glass cover slips coated with poly-(D)-lysine hydrobromide (Sigma-Aldrich, 0.1 mg/ml in H_2O) in DMEM, FBS (10%), B, H with or without doxycycline (10 ng/ml final conc.). Cells were incubated with 200 μl of DMEM, FBS (10%) containing 2 μl NucBlue™ Live ReadyProbes™ Reagent Hoechst 33342 (ThermoFisher Scientific) and *O*-acetylated benzylguanine fluorescein isothiocyanate (ac-BG-FITC, 2 μM final conc.). After fixation with formaldehyde (3.7% final conc.) and permeabilization with Triton X-100 (Carl Roth, #3051.3, 0.1% final conc.), cover slips were mounted on microscope slides using Dako Fluorescence Mounting Medium (AgilentDako) and dried over-night at 4°C. Images were taken with an AXIO Observer.Z1 (Zeiss), Colibri.2 light source under 63x magnification.

Functional CTNNB1 assay

Editase expressing cells were transfected as described above in technical duplicates (SNAP-CDAR-S) or quadruplicates (Cas RESCUE-S). Cells were transfected with either M50 Super 8x TOPFlash (Addgene, #12456) or M51 Super 8x FOPFlash (Addgene, #12457) and with pcDNA3.1 expressing Renilla Luciferase for normalization. Samples were also transfected either empty, or with CTNNB1 T41-targeting $(snap)_2$-gRNA or gRNA expressing plasmid (RESCUE-S), or with PPIB R7C-targeting $(snap)_2$-gRNA or gRNA expressing plasmid (RESCUE-S), or with CTNNB1 T41-targeting NH_2-gRNA or plasmid expressing no gRNA (RESCUE-S) in a 96-well format. Cells were lysed with 30 μl (SNAP-CDAR-S) or 20 μl (Cas RESCUE-S) per well of Passive Lysis Buffer (Promega) and shaken for 15 min. at room temperature. Luciferase signal was measured as described before (26) using the Dual-Luciferase® Reporter Assay System (Promega) following manufacturer's instructions with a Spark 10 M plate reader (Tecan). Briefly, 10 μl of each replicate (two of the four wells for RESCUE-S were pooled) were measured by addition of 35 μl of Luciferase Assay Reagent II (LAR II, Promega) and 35 μl of Stop & Glo® Reagent, and measured for 10 seconds, respectively. Editing yield was determined as described above. All luminescence measurements and editing yield determinations were conducted in biological triplicates.

Editing of STAT3 pS727

3×10^5 Flp-In T-Rex cells expressing SNAP-CDAR-S were seeded in a 24-well format in DMEM, FBS (10%), D (10 ng/ml) to induce transgene expression. After 24 h, 3.2×10^5 cells were transfected with 20 pmol of STAT3 S727F-targeting NH_2-gRNA, $(snap)_2$-gRNA, or PPIB R7C-targeting $(snap)_2$-gRNA (quadruple of 96-well format) using 2 μl Lipofectamine™ RNAiMAX (invitrogen) per transfection. This transfection was repeated on day 2, 4 and 6 after the first transfection. Endpoint was at day 8. A fraction of cells was used for RNA isolation and editing analysis as described above. The rest of the cells was lysed with RIPA lysis and extraction buffer (ThermoFisher Scientific) supplemented with cOmplete™ Tablets, Mini, EDTA-free *EASY*pack Protease Inhibitor Cocktail (Roche) and PhosStop *EASY*pack (Roche). Protein concentration was determined by Pierce™ BCA Protein Assay Kit (ThermFisher Scientific) in a Tecan Plate Reader. 30 μg of total protein was run on a Novex™ WedgeWell™ 8 to 16%, Tris-glycine, 1.0 mm, Mini Protein Gel (ThermoFisher Scientific) with 200 V for 60 min. Blotting was performed with a Mini Trans-Blot Cell® (Bio-Rad) at 100 V for 60 min. For protein detection, membranes were incubated with monoclonal anti-β-actin antibody produced in mouse (Sigma, 1:5000 dil.) and either Stat3 (DRZ2G) Rabbit mAb (CellSignaling, 1:1000 dil.) or P-Stat3 (S727) (D8C2Z) Rabbit mAb (CellSignaling, 1:1000 dil.). Membranes were then incubated with peroxidase AffiniPure goat anti-mouse IgG (Jackson Immuno Research, #115-035-003, 1:10000 dil.), and peroxidase AffiniPure goat anti-rabbit IgG (Jackson Immuno Research, #111-035-003, 1:10000 dil.). Images were taken with Odyssey FC Imager (Li-Cor® Biosciences).

Next generation sequencing

Cells expressing SNAP-CDAR-S or Cas RESCUE-S were transfected with or without $(snap)_2$-gRNA (2.5 pmol) or gRNA expression vector (300 ng), respectively, targeting PPIB R7C in technical duplicates under constant Dox (10 ng/ml) induction. Total RNA was

isolated with RNeasy MinElute Cleanup Kit (Qiagen) following manufacturer's instructions. After DNase I (NewEngland BioLabs) digest, samples were again purified with RNeasy MinElute Cleanup Kit. Editing yields were first determined by Sanger Sequencing, as described above. Next Generation Sequencing was performed by CeGaT with a NovaSeq6000 (Illumina) to generate 50 Mio. 2 × 100 bp paired-end reads per sample. Library was prepared with TruSeq Stranded mRNA Kit (Illumina). Lanes of raw RNA sequencing data of same samples were pulled together, and adapter sequences were trimmed with Trim Galore (v0.6.5; http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Sequencing alignment to the human reference genome (hg19) was performed using STAR (v. 2.7.10a). hg19 and the RefSeq annotation are publicly available at the genome browser at UCSC. For alignment uniquely mapped (STAR option: –outFilterMultimapNmax 1) reads were considered to prevent multimapping of regions of high similarity. Next, data (bam files) were deduplicated, sorted, and indexed using SAMtools (v1.9; http://samtools.sourceforge.net). SNVs calling was performed with REDItools (v2; https://github.com/tflati/reditools2.0). Developers' recommendations were taken into consideration for preceding data preparation. As previously performed, only high-quality sites (min. MeanQ > 30 in REDItools2) were considered. Editing site were called when well-covered i.e. minimum 50 reads (in summary of duplicates per sample), showing ≥5% editing frequency compared to the control. For sites matching criteria, fisher's exact test was performed. Significance was defined for all samples showing an adjusted *P*-value < 0.01. Genomic coordinates were annotated with Variant Effect Predictor (VEP; v102) using the following command line:

vep –input file input..txt –fasta reference.fa –output_file output.txt –species homo_sapiens –tab –cache –dir_cache ../Human/dir –no_check_variants_order –transcript_version –canonical –ccds –hgvs –symbol –gene_phenotype –pubmed –variant_class –pick –offline –force_overwrite

Further information

For more detailed protocols and guide RNA sequences, please see additional Supplementary Information and additional Supplementary files. Detailed information on reagents, enzymes, antibodies, and kits as well as cell lines used in this study are presented in the Supporting Information.

RESULTS

The APOBEC1-SNAP tool suffers from low C-to-U editing yields and limited programmability

Recently, we demonstrated the harnessing of the murine APOBEC1 deaminase for site-directed C-to-U RNA base editing. For this, the SNAP-tag was fused to the C-terminus of the full length APOBEC1 enzyme, resulting in an editor called APO1S (Figure 1A) (20). In transgenic cell lines that co-express APO1S together with SNAP-ADAR1Q, moderate editing yields were achieved on an eGFP reporter gene, but editing yields on the endogenous GAPDH transcript stayed below 20%. By targeting the eGFP reporter, we now tried several means to improve editing yields. On the guide RNA side, the positioning of the guide RNA four to five nucleotides upstream with respect to the target cytidine was most important (Figure 1B, C). On the protein side, the localization of the editing enzyme to the cytosol was particularly necessary (Figure 1D, E, Supplementary Figures S1-S6). Nevertheless, the APO1S editor suffered overall from low editing yields on endogenous targets and from low programmability (Figure 1F), meaning that transfer to endogenous transcript was particularly difficult followed by notable guide RNA-dependent bystander editing when on-target editing was successful (Supplementary Figure S7).

The SNAP-CDAR-S tool combines high editing yields with excellent programmability

In contrast, the Cas-13-mediated C-to-U editing tool called RESCUE applies a C-to-U deaminase that was evolved from the A-to-I deaminase ADAR2 (24), and shares with ADAR2 its strong substrate preference for double-stranded RNA. Indeed, the RESCUE tool seems to have much better programmability and on-target editing was reliably obtained when the target site was positioned inside the guide RNA / mRNA duplex. However, Cas13-mediated C-to-U editing suffers from global and local C-to-U and A-to-I off-target editing, and attempts to create more precise tools, like Cas13-RESCUE-S, came along with largely reduced on target editing yields, hardly above 10% on endogenous transcripts (24). However, we were wondering how the engineered cytidine deaminase acting on RNA (CDAR) domain would work in the context of a SNAP-tagged fusion protein (12). For this, we fused the evolved, high-fidelity deaminase of the RESCUE-S tool to the C-terminus of a SNAP-tag (11,14) to obtain the SNAP-CDAR-S tool (Figure 2A). We stably integrated a single copy of the SNAP-CDAR-S transgene into HEK 293 cells by using the Flp-In approach, (14,19) and found homogenous transgene expression under control of doxycycline. Similar to the closely related A-to-I editing enzyme SNAP-ADAR2Q (Supplementary Figure S1), the SNAP-CDAR-S tool was mainly localized to the cytosol (Figure 2B).

Our initial guide RNA design was inspired from our experience with the SNAP-ADAR tool and was tested for the editing of a 5′-ACA codon in a co-transfected eGFP reporter transcript. Initial guide RNAs were 22 nt long, chemically modified by 2′-O-methylation outside the target base triplet (27), which is the targeted cytidine plus its two closest neighboring bases, and carried a $(snap)_2$ self-labeling moiety (20) at the 5′-end for the recruitment of two SNAP-CDAR-S effectors per guide RNA. The exact composition, sequence and chemistry, of all guide RNAs can be found in the Supplementary Information (Supplementary Table T1a). While the SNAP-ADAR tool prefers a relatively central positioning of the target nucleobase, the SNAP-CDAR-S effector gave clearly better yields when the target cytidine was located near the 5′-terminus of the guide RNA, e.g. design 3-C-18 in Figure 2C, in good agreement with data from the Cas13-RESCUE (24) tool. Next, we took a closer look at the guide RNA design for the editing of the endogenous

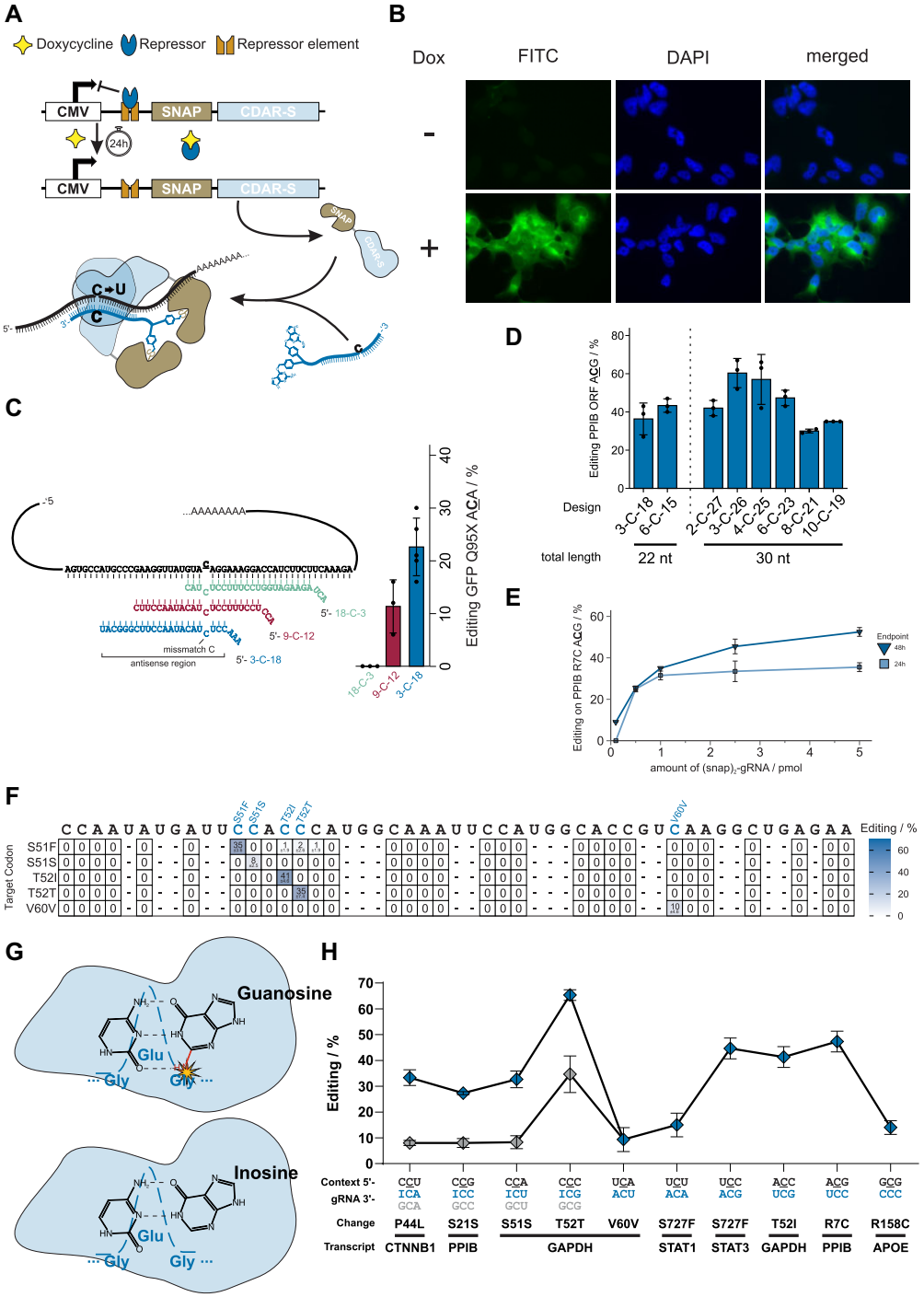


Figure 2. Guide RNA design and performance of the SNAP-CDAR-S tool. (**A**) Scheme of the doxycycline-inducible SNAP-CDAR-S tool and the guide RNA-dependent editing reaction. (**B**) Analysis of SNAP-CDAR-S transgene expression and localization by fluorescence microscopy. SNAP-CDAR-S was stained with SNAP-tag-reactive BG-FITC (green channel). (**C**) Dependency of editing yield of the Q95X site (5′-ACA) in the eGFP reporter gene on the positioning of the guide RNA (all 22 nucleotides length, 2′OMe modification on all nucleotides except for mismatch C and flanking nucleotides) relative to the target site (symmetric versus asymmetric). (**D**) Fine-tuning of guide RNA length (22, 30 nt) and positioning (of mismatch cytidine C) for optimal editing performance. (**E**) Dependency of RNA editing yield on the amount of transfected guide RNA (pmol/96 well). (**F**) Programmability and precision of the SNAP-CDAR-S tool. Five guide RNAs (each 6-C-23, 2′OMe gapmer design) against five nearby cytidine sites on the endogenous GAPDH transcript, each with a distinct codon context, were applied and the on-target and the C-to-U and A-to-I bystander off-target editing was determined by Sanger sequencing. (**G**) Scheme explaining the positive effect of inosine in guide RNAs for targeting 5′-CCN codons. Pairing of the 5′ cytidine in a 5′-CCN context with a guanosine leads to a steric clash of CDAR´s glycine with the guanosine's exocyclic NH_2-group (28). Inosine lacks that NH_2-group thus avoiding steric clash. (**H**) Codon scope of the SNAP-CDAR-S tool and effect of inosine in 5′-CCN codons. Given are editing yields for various codons on various endogenous transcripts when applying non-optimized guide RNAs of the standard design (6-C-23, 2′OMe gapmer). For each of the four 5′-CCN codons (N = U, C, A, G), the editing yields of guide RNAs are compared that contained either an inosine (I) or a guanosine (G) opposite the cytidine preceding the on-target cytidine. Data in (C), (D), (E), (F) and (G) are shown as the mean ± s.d. of $N \geq 2$ independent experiments.

PPIB transcript, specifically, by targeting a 5′-ACG codon in its coding region (ORF). Here, we varied the length of the guide RNA (22 nt and 30 nt) and the positioning of the guide RNA relative to the target cytidine, see Figure 2D. With 60% editing yield, we found the best performing guide RNA to be 30 nt long, positioning the targeted cytidine close to the 5′-end (position 4, 3-C-26) of the guide RNA in the substrate duplex. However, also other guide RNA designs gave good editing yields and the optimal design might vary to some extent for each target sequence, as suggested for the original Cas13-RESCUE-S tool (24). As we sought to determine a universal guideRNA design, we compared editing yields on further endogenous transcripts putting the target C in position 4 (3-C-26) or position 7 (6-C-23, Supplementary Figure S8). In contrast to the target site on PPIB, positioning target C at position 7 showed substantially higher editing yields for all selected sites (Supplementary Figure S8). In addition, we transferred designs previously reported (24) ideal for four endogenous sites to our SNAP-CDAR-S tool and compared them to our 6-C-23 standard designs (Supplementary Figure S9). For all targets, our standard design gave similar or even better editing yield then the previously described ideal design. This indicated that 6-C-23 can be considered a universal design for the SNAP-CDAR-S tool. We then continued to analyze the performance further. Editing yields were saturating when ≥2.5 pmol/96 well (20 nM) guide RNA were transfected (Figure 2E). To characterize the scope, programmability and precision of the SNAP-CDAR-S tool, we targeted five different guide RNAs (all 30 nt, 6-C-23, and 2′-OMe) against five different cytidine bases, which were all located in close proximity in the ORF of the endogenous GAPDH transcript and determined on-target as well as C-to-U and A-to-I bystander editing yields. We found excellent programmability, with good on-target yields (8% to 41%) and with lacking bystander editing (detection limit Sanger sequencing ca. 5%) at neighboring cytidine or adenosine bases (Figure 2F, the same was found for an alternative 3-C-18 guide RNA design, see Supplementary Figure S10). 2′-O-Methylation was shown in the past to block bystander A-to-I editing very efficiently in SNAP-ADAR tools (14,27), and this may contribute here to the high precision of the targeted editing too. However, we were not fully satisfied with the editing yield at the 5′-CCA codon (GAPDH S51S), which achieved only 8% with the best design (6-C-23, a 3-C-18 design gave < 5%). This limited scope was also reported for the Cas13-based RESCUE tool (24) and resembles the codon preference of the ancestor ADAR2 (25) protein. A recent structural analysis of the ADAR2 deaminase bound to a dsRNA substrate revealed a steric clash between the peptide backbone of glycine 489 and the minor groove face of G = C base pairs residing at the 5′ neighboring position to the target adenosine (28). This steric clash could be relaxed by replacing the 5′-neighboring G = C base pair with sterically less demanding I = C base pair (lacking an exocyclic amino group), simply by pairing the 5′-CCA target codon with a 5′-UCI sequence in the guide RNA (Figure 2G). Indeed, we found a 3-fold improved editing yield of 32% for the respective site in GAPDH (Figure 2H, Supplementary Figure S10B). We then systematically tested the principle for all four potential 5′-CCN codons (N = A, U, G, C) and found that an inosine base opposite the 5′-neighboring cytidine always improved editing at the targeted cytidine base (Figure 2H). Even for the well-edited 5′-CCC codon (34%), we could still achieve a notable gain in editing yield (66%).

SNAP-CDAR-S clearly outperforms Cas13-RESCUE-S on endogenous targets

To benchmark the SNAP-tagged tool with the Cas13-based tool, we generated an analogous 293 Flp-In T-REx cell line stably expressing the Cas13-based RESCUE-S on doxycycline induction. In the original work (24), RESCUE-S has always been applied by means of transient overexpression, however, this often leads to high variability in editing yields and artifacts in off-target analyses (12). We tested both editing tools side-by-side for the editing of eight different sites on five different endogenous transcripts (GAPDH, PPIB, CTNNB1, STAT3, STAT1) and one disease-relevant cDNA (APOE). Most target sites were taken from the original Cas13-RESCUE-S publication (24) so that optimal Cas guide RNAs have already been reported for each of them (Supplementary Table T1b). We repeated these experiments by transfecting 300 ng/96 well of the plasmid-borne optimal guide RNAs into the stable Cas13-RESCUE-S cells. The guide RNAs for the SNAP-CDAR-S cell lines were not optimized, but we simply transfected 5 pmol/96 well chemically stabilized, 30 nt guide RNAs of the 6-C-23 standard design. Nevertheless, the SNAP-CDAR-S tool clearly outcompeted the Cas13-RESCUE-S tool on all eight targets, achieving editing yields between 10% and 50% (Figure 3A), while the editing yields of the Cas13-RESCUE-S tool did not achieve editing yields above 10%, in accordance with the original report (24). For four targets, only SNAP-CDAR-S, but not Cas13-RESCUE-S, was able to achieve detectable editing (PPIB S21G, CTNNB1 H63Y and P44L, STAT1 S727F). Interestingly, three out of these four examples target 5′-CCN codons, which is readily done by the SNAP-CDAR-S approach with inosine-containing guide RNAs, highlighting that the SNAP-CDAR-S approach does not only give generally higher editing yields but also increases the codon scope towards 5′-CCN sites.

Several of the indicated targets are of clinical interest. The removal of threonine 41 from the β-catenin protein inactivates a degron and thus stabilizes the protein. (29) Enhanced levels of β-catenin could be applied to boost liver regeneration or wound healing transiently (30). We benchmarked the activation of the Wnt pathway by the SNAP-CDAR-S versus its analog Cas13 tool in a plasmid-borne luciferase assay, following a protocol reported before (24). While the SNAP-tagged tool achieved 21% editing yield and an 8-fold increase in β-catenin activity (Figure 3B, Supplementary Figure S11), the Cas13-RESCUE-S tool gave only 7% editing yield and 1.6-fold increase. The STAT3 protein (signal transducer and activator of transcription 3) is a multifunctional signaling molecule, which acts as a transcription factor in the nucleus, or translocates to the mitochondrium, and modulates immune response and metabolism. Its hyperactivation plays an important role in autoimmune disease, sterile inflammation, and cancer (31). Here, we

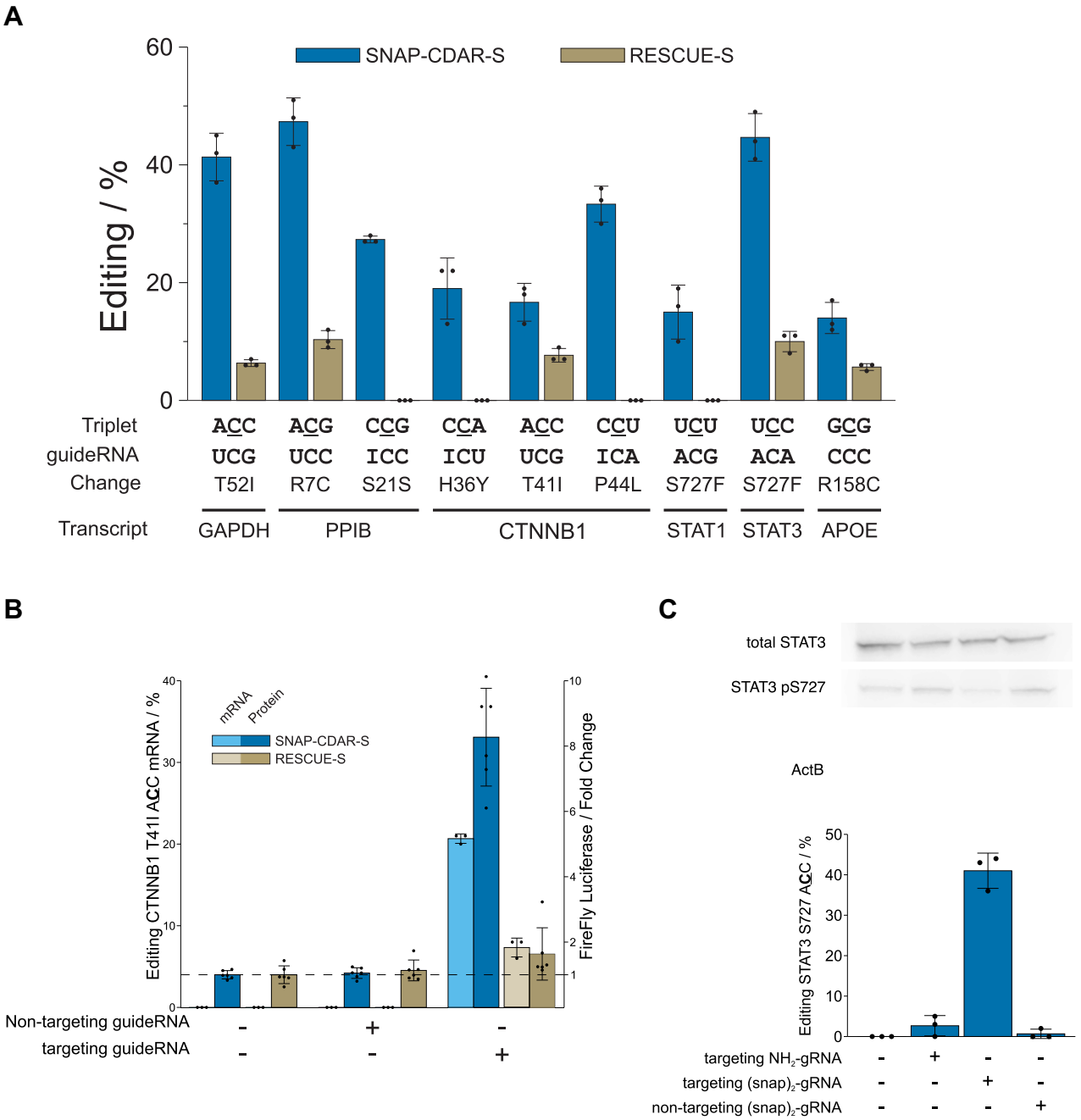


Figure 3. Benchmark with Cas13-RESCUE-S on endogenous targets and applications. (**A**) Comparison of editing yields at various sites on various endogenous targets and one disease-relevant cDNA (APOE) comparing SNAP-CDAR-S with standard guide RNAs (30 nt, 6-C-23, 2′OMe gapmer) versus Cas13 RESCUE-S with plasmid-borne optimized Cas guide RNAs. Both editing enzymes were expressed from the same single genomic locus. (**B**) Comparing both tools, SNAP-CADR-S versus Cas RESCUE-S, for the activation of β-catenin by RNA editing. Given are C-to-U editing yields (T41I) and the luminescence-based read-out of pathway activation. For further controls, see Supplementary Figure S11. (**C**) Editing of the regulatory phospho-site serine 727-to-glycine in STAT3, read-out of editing yield by Sanger sequencing, and amount of total STAT3 and pS727 STAT3 protein by western blot. Data in (A), (B) and (C) are shown as the mean ± s.d. of $N = 3$ independent experiments.

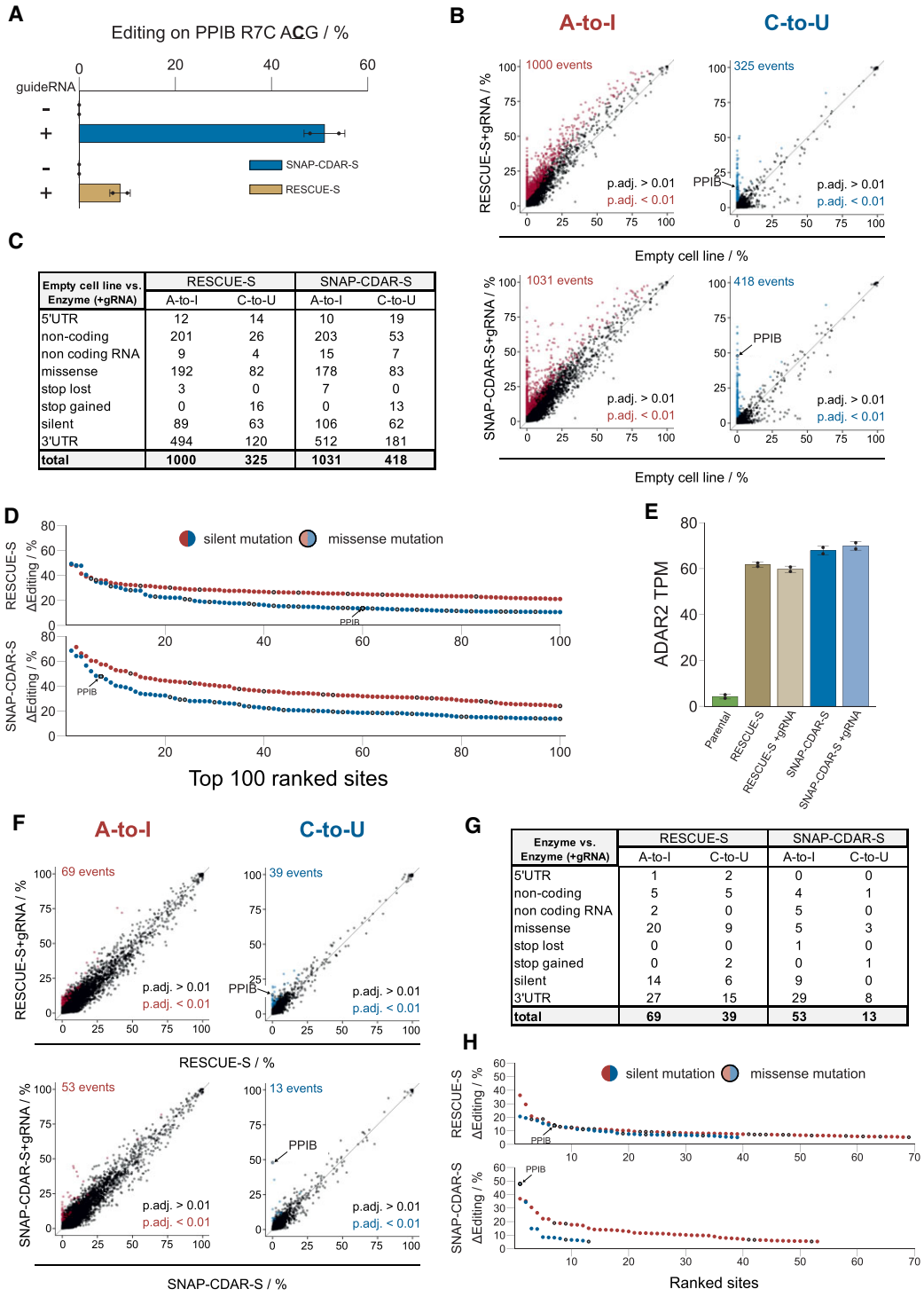
removed serine 727 from STAT3, a functionally important phosporylation site. We could achieve up to 41% serine-to-glycine editing, which was accompanied by a visible reduction in S727 phosphorylation as determined by Western blot (Figure 3C, Supplementary Figure S12). Finally, we aimed at introducing a protective genotype into the apolipoprotein E (APOE) transcript, introducing the rs7412 SNP (R158C), which could transfer the neutral ε3 allel (ca. 78% caucasian carriers) into the protective ε2 allel, which was shown to largely reduce the risk for atherosclerosis (32). However, given the non-preferred nature of the codon (5′-GCG), and the very high GC content of the surrounding sequence space, an editing yield of only 14% was achieved, still clearly outcompeting the Cas13 tool (Figure 3A).

Both tools show moderate global A-to-I and C-to-U off-target editing

We used next-generation sequencing of the poly(A)+ transcriptome (10 GB per condition) to assess the transcriptome-wide A-to-I and C-to-U off-target editing of the SNAP-CDAR-S versus its analog Cas13-RESCUE-S. We took RNA from cells expressing the respective editing effector in the presence and absence of the respective guide RNA and compared them to Flp-In T-REx cells not expressing an engineered effector (14). First, we compared the editing reactions against the empty Flp-In T-REx cell line and were able to detect the on-target editing event (PPIB Arg7Cys) with editing yields of 48% for SNAP-CDAR-S and 14% for Cas13-RESCUE-S, which were confirmed by Sanger sequencing (Figure 4A). Beside the on-target editing, we found around 1000 A-to-I and three to four hundred C-to-U off-target events for both effectors (Figure 4B). As seen for ADAR-based effectors before, (10,12,33) A-to-I off-target editing was a combination of enhanced editing at known sites and editing at novel sites, whereas the large majority of C-to-U off-target editing were novel sites. We further analyzed the outcomes of the editing reactions and found that only a moderate number of all editing events resulted in missense mutations (Figure 4C). At less than ten missense sites, the change in the off-target editing yield was increased above 25%, indicating that most missense sites are only marginally affected (Figure 4D). The patterns between the two effectors were very similar, which was expected given that the CDAR domain of both editing tools is identical. The presence of the editing tools gave no larger changes in gene expression (Supplementary Figure S13), and both editing effectors were expressed to similar TPM levels (Figure 4e). Finally, we analyzed the guide RNA-dependent changes in editing. Clearly, the vast majority of off-target editing came from the presence of the editing enzymes and was guide RNA-independent (Figure 4F, G). The guide RNA-dependent C-to-U editing was very clean for the SNAP-CDAR-S tool. In contrast, the on-target editing with Cas13-RESCUE-S was covered by a small number of further editing events. This became also visible when we plotted all guide RNA-dependent changes in editing yields (Figure 4H). While the on-target site gave the largest ΔEditing value for the SNAP-CDAR-S tool, the Cas13-RESCUE-S tool gave six A-to-I and another six C-to-U off-target events with higher change in editing level. Overall, both enzymes were expressed to comparable TPM levels, gave very similar patterns and levels of global off-target editing and mainly differed in the 5-fold higher on-target editing yield of the SNAP-CDAR-S tool.

DISCUSSION

In comparison to the APOBEC1 enzyme, the CDAR domain, evolved from ADAR2, performs considerably better in targeted RNA base editing tools. While the CDAR domain was taken from the Cas13-mediated RESCUE approach (24), we could show here that this deaminase domain works particularly well when the self-labeling SNAP-tag is applied as the RNA-targeting mechanism. Compared to the Cas13-RESCUE-S, the SNAP-CDAR-S gave reliably higher on-target yields with less bystander editing, while the global A-to-I and C-to-U off-target effects were comparable. A reason for the superior efficiency of the SNAP-tagged tool might be the chemical design and the covalent bond that tethers the guide RNA to the SNAP-tag and may foster the encounter of guide RNA, target RNA and editing effector (12,14). Regarding the design of the guide RNAs with respect to chemical modifications, we found that lessons learned from the engineering of the closely related SNAP-ADAR tool (14,27) could be largely transferred. In particular, the general guide RNA design with 2′-O-methylation at the ribose moieties outside the base triplet and the usage of non-encodable bases like inosine opposite 5′-C<u>C</u>N codons have contributed to the improved performance so that editing yields between 10% and 50% are regularly achieved in 5′-H<u>C</u>N (H = C, A, U) codons, and only 5′-G<u>C</u>N codons remain challenging. To our knowledge, our data is the first report of stable integration of the Cas13-RESCUE-S tool and shows that it functions as well under genomic integration as it does via plasmid overexpression. Even though Cas13-RESCUE-S was presented as a high-fidelity enzyme with largely reduced global A-to-I off-target editing before (24), there still remains notable A-to-I as well as C-to-U off-target editing, which might have been underestimated in the prior study, where off-target analysis was done on reporter cDNA under co-transfection of the editing enzyme. Our work further allows to compare the A-to-I off-target effects of the SNAP-CDAR-S deaminase directly to the related, stably integrated wildtype and hyperactive (E488Q) mutant of SNAP-ADAR2 (14). While the off-target A-to-I editing of the SNAP-CDAR-S tool is clearly below that of the hyperactive, off-target-prone SNAP-ADAR2 E488Q mutant (14), the SNAP-CDAR-S tool still has notably frequent off-target A-to-I editing when compared to the wildtype SNAP-ADAR2 enzyme (see Supplementary Figure S14), which is clearly not yet optimal for a C-to-U editing enzyme and may require further engineering efforts to generate a pure C-to-U editing enzyme. Recently, improved C-to-U editing tools based on the Cas13-RESCUE platform have been described (34,35). Engineering efforts allowed to largely reduce the size of the Cas protein, e.g. to enable AAV-mediated delivery. However, these tools remain built on the original C-to-U deaminase domain taken from the RESCUE(-S) tool so that global C-to-U and in particular A-to-I off-target editing remains observable. In contrast, a split version of the CDAR domain has recently been described to strongly improve editing precision, e.g. by largely abolishing such global off-target editing. (36) Specifically, the tool uses an orthogonal trans-tethering approach, steering one CDAR half via an MS2/MCP and the other half via a λN/BoxB interaction to the target mRNA. While this was working in principle, the C-to-U editing efficiency remained very low (e.g. around 5%). Finally, C-to-U editing tools have been constructed based on the RNA/DNA editing enzyme APOBEC3A, including the Cas13-based tool CURE (22) and the PUF domain-based tool REWIRE (37). While both tools enable programmable editing, they also come with specific limitations. These include global off-target C-to-U editing at the RNA and DNA level, but also a strongly limited codon scope, e.g. 5′-U<u>C</u>R (R = A, G). Overall, the SNAP-CDAR-S tool adds a reliable and efficient enzyme with large codon scope to the tool box for



Empty cell line vs. Enzyme (+gRNA)	RESCUE-S		SNAP-CDAR-S	
	A-to-I	C-to-U	A-to-I	C-to-U
5'UTR	12	14	10	19
non-coding	201	26	203	53
non coding RNA	9	4	15	7
missense	192	82	178	83
stop lost	3	0	7	0
stop gained	0	16	0	13
silent	89	63	106	62
3'UTR	494	120	512	181
total	**1000**	**325**	**1031**	**418**

Enzyme vs. Enzyme (+gRNA)	RESCUE-S		SNAP-CDAR-S	
	A-to-I	C-to-U	A-to-I	C-to-U
5'UTR	1	2	0	0
non-coding	5	5	4	1
non coding RNA	2	0	5	0
missense	20	9	5	3
stop lost	0	0	1	0
stop gained	0	2	0	1
silent	14	6	9	0
3'UTR	27	15	29	8
total	**69**	**39**	**53**	**13**

Figure 4. Precision of SNAP-CDAR-S versus Cas13-RESCUE-S as determined by NGS. (**A**) Editing yields as determined by Sanger sequencing prior to NGS. (**B**) Plot of total off-target events of the respective effector + guide RNA against the empty cell line. Significantly differently edited sites (adjusted *P*-value < 0.01) are colored in red (A-to-I) and blue (C-to-U), respectively. (**C**) Total off-target effects sorted by categories. (**D**) Changes in editing yields (ΔEditing) of the 100 top-ranked editing sites color-coded for A-to-I (red) and C-to-U (blue) editing. The on-target editing event is marked by an arrow. Missense and silent mutations are indicated by empty and filled symbols, respectively. (**E**) Transgene expression levels as determined by TPM values of ADAR2 (ADARB1). Given that CDAR-S is an ADAR2 deaminase mutant, both transgenes (SNAP-CDAR-S, Cas13-RESCUE-S) were annotated as such. ADAR2 itself is not expressed in the parental 293 cell line. Both transgenes were expressed to comparable TPM levels and did not change upon presence of the guide RNA. (**F**) Plot of guide RNA-dependent off-target events of the respective effector + guide RNA against the respective effector. Significantly differently edited sites (adjusted *P*-value < 0.01) are colored in red (A-to-I) and blue (C-to-U), respectively. (**G**) Guide RNA-dependent off-target effects sorted by categories. (**H**) Changes in editing yields (ΔEditing) of the top-ranked editing sites, color-coded for A-to-I (red) and C-to-U (blue) editing. The on-target editing event is marked by an arrow. Missense and silent mutations are indicated by empty and filled symbols, respectively. Significance in panels (B) and (F) was tested by Fisher's exact test (two-sided), $n = 2$ independent experiments.

targeted RNA base editing. It may also be worth mentioning that the SNAP-CDAR-S tool avoids protein parts taken from bacterial origin, like Cas proteins, which raise concerns regarding immunogenicity upon their perpetual expression, (38) which would be required in many therapeutic settings. Even though editing efficiency and precision of SNAP-CDAR-S are not yet perfect, the tool clearly outperforms the original Cas13-based RESCUE-S and lays a basis for further engineering in the future.

DATA AVAILABILITY

The raw NGS data is uploaded on the NCBI BioProject Server under ID PRJNA948645.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to express our gratitude to Dr Nina Papavasiliou and Dr Rafail Tasakis for their expertise and support on the analysis of the Next Generation Sequencing data. We would further like to express our gratitude to Carolin Fruhner for the superb technical expertise she provided without hesitation.

FUNDING

Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) [430214260, STA1053/7-1 to T.S.]; DFG priority program SPP 1784 [404867268 to T.S.]; DFG TRR319 RMaP [439669440 project A4 to S.G.]; European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program [647328 to T.S.]. Funding for open access charge: University of Tübingen; DFG.

Conflict of interest statement. T.S. and N.L. hold patents on site-directed RNA editing. T.S. is founder, share holder of and consultant to AIRNA Corporation.

REFERENCES

1. Pecori,R., Di Giorgio,S., Lorenzo,J.P. and Papavasiliou,F.N. (2022) Functions and consequences of AID/APOBEC-mediated DNA and RNA deamination. *Nat. Rev. Genet.*, **23**, 505–518.
2. Teng,B., Burant,C.F. and Davidson,N.O. (1993) Molecular cloning of an apolipoprotein B messenger RNA editing protein. *Science*, **260**, 1816–1819.
3. Powell,L.M., Wallis,S.C., Pease,R.J., Edwards,Y.H., Knott,T.J. and Scott,J. (1987) A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine. *Cell*, **50**, 831–840.
4. Yamanaka,S., Balestra,M.E., Ferrell,L.D., Fan,J., Arnold,K.S., Taylor,S., Taylor,J.M. and Innerarity,T.L. (1995) Apolipoprotein B mRNA-editing protein induces hepatocellular carcinoma and dysplasia in transgenic animals. *Proc. Natl Acad. Sci. U.S.A.*, **92**, 8483–8487.
5. Christofi,T. and Zaravinos,A. (2019) RNA editing in the forefront of epitranscriptomics and human health. *J. Transl. Med.*, **17**, 319.
6. Fossat,N., Tourle,K., Radziewic,T., Barratt,K., Liebhold,D., Studdert,J.B., Power.,M., Jones,V., Loebel,D.A. and Tam,P.P. (2014) C to U RNA editing mediated by APOBEC1 requires RNA-binding protein RBM47. *EMBO Rep.*, **15**, 903–910.
7. Mehta,A., Kinter,M.T., Sherman,N.E. and Driscoll,D.M. (2000) Molecular cloning of apobec-1 complementation factor, a novel RNA-binding protein involved in the editing of apolipoprotein B mRNA. *Mol. Cell. Biol.*, **20**, 1846–1854.
8. Richardson,N., Navaratnam,N. and Scott,J. (1998) Secondary structure for the apolipoprotein B mRNA editing site. *J. Biol. Chem.*, **273**, 31707–31717.
9. Sharma,S. and Baysal,B.E. (2017) Stem–loop structure preference for site-specific RNA editing by APOBEC3A and APOBEC3G. *PeerJ*, **5**, e4136.
10. Rees,H.A. and Liu,D.R. (2018) Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.*, **19**, 770–788.
11. Keppler,A., Gendreizig,S., Gronemeyer,T., Pick,H., Vogel,H. and Johnsson,K. (2003) A general method for the covalent labeling of fusion proteins with small molecules *in vivo*. *Nat. Biotechnol.*, **21**, 86–89.
12. Vogel,P. and Stafforst,T. (2019) Critical review on engineering deaminases for site-directed RNA editing. *Curr. Opin. Biotechnol.*, **55**, 74–80.
13. Stafforst,T. and Schneider,M.F. (2012) An RNA-deaminase conjugate selectively repairs point mutations. *Angew. Chem. Int. Ed.*, **51**, 11166–11169.
14. Vogel,P., Moschref,M., Li,Q., Merkle,T., Selvasaravanan,K.D., Li,J.B. and Stafforst,T. (2018) Efficient and precise editing of endogenous transcripts with SNAP-tagged ADARs. *Nat. Methods*, **15**, 535–538.
15. Cox,D.B.T., Gootenberg,J.S., Abudayyeh,O.O., Franklin,B., Kellner,M.J., Joung,J. and Zhang,F. (2017) RNA editing with CRISPR-Cas13. *Science*, **358**, 1019–1027.
16. Montiel-Gonzalez,M.F., Vallecillo-Viejo,I.G., Yudowski,A. and Rosenthal,J.J.C. (2013) Correction of mutations within the cystic fibrosis transmembrane conductance regulator by site-directed RNA Editing. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 18285–18290.
17. Stroppel,A.S., Lappalainen,R. and Stafforst,T. (2021) Controlling site-directed RNA editing by chemically induced dimerization. *Chemistry EurJ*, **27**, 12300–12304.
18. Hanswillemenke,A., Kuzdere,T., Vogel,P., Jékely,G. and Stafforst,T. (2015) 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.
19. Vogel,P., Hanswillemenke,A. and Stafforst,T. (2017) Switching protein localization by site-directed RNA editing under control of light. *ACS Synth. Biol.*, **6**, 1642–1649.
20. Stroppel,A.S., Latifi,N., Hanswillemenke,A., Tasakis,R.N., Papavasiliou,F.N. and Stafforst,T. (2021) Harnessing self-labeling enzymes for selective and concurrent A-to-I and C-to-U RNA base editing. *Nucleic Acids Res.*, **49**, e95.
21. Bhakta,S., Sakari,M. and Tsukahara,T. (2020) RNA editing of BFP, a point mutant of GFP, using artificial APOBEC1 deaminase to restore the genetic code. *Sci. Rep.*, **10**, 17304.
22. Huang,X., Lv,J., Li,Y., Mao,S., Li,Z., Jing,Z., Sun,Y., Zhang,X., Shen,S., Wang,X. *et al.* (2021) Programmable C-to-U RNA editing using the human APOBEC3A deaminase. *EMBO J.*, **40**, e108209.
23. Kuttan,A. and Bass,B.L. (2012) Mechanistic insights into editing-site specificity of ADARs. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, E3295–E3304.
24. Abudayyeh,O.O., Gootenberg,J.S., Franklin,B., Koob,J., Kellner,M.J., Ladha,A., Joung,J., Kirchgatterer,P., Cox,D.B.T. and Zhang,F. (2019) A cytosine deaminase for programmable single-base RNA editing. *Science*, **365**, 382–386.
25. Eggington,J.M., Greene,T. and Bass,B.L. (2011) Predicting sites of ADAR editing in double-stranded RNA. *Nat. Commun.*, **2**, 319–319.
26. Reautschnig,P., Wahn,N., Wettengel,J., Schulz,A.E., Latifi,N., Vogel,P., Kang,T., Pfeiffer,L.S., Zarges,C., Naumann,U. *et al.* (2022) CLUSTER guide RNAs enable precise and efficient RNA editing with endogenous ADAR enzymes in vivo. *Nat. Biotechnol.*, **40**, 759–768.
27. Vogel,P., Schneider,M.F., Wettengel,J. and Stafforst,T. (2014) Improving site-directed RNA editing in vitro and in cell culture by chemical modification of the GuideRNA. *Angew. Chem.*, **126**, 6382–6386.
28. Matthews,M.M., Thomas,J.M., Zheng,Y., Tran,K., Phelps,K.J., Scott,A.I., Havel,J., Fisher,A.J. and Beal,P.A. (2016) Structures of human ADAR2 bound to dsRNA reveal base-flipping mechanism and basis for site selectivity. *Nat. Struct. Mol. Biol.*, **23**, 426–433.

29. Morin,P.J., Sparks,A.B., Korinek,V., Barker,N., Clevers,H., Vogelstein,B. and Kinzler,K.W. (1997) Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC. *Science*, **275**, 1787–1790.
30. Zhao,L., Jin,Y., Donahue,K., Tsui,M., Fish,M., Logan,C.Y., Wang,B. and Nusse,R. (2019) Tissue repair in the mouse liver following acute carbon tetrachloride depends on injury-induced Wnt/β-catenin signaling. *Hepatology*, **69**, 2623–2635.
31. Huynh,J., Chand,A., Gough,D. and Ernst,M. (2019) Therapeutically exploiting STAT3 activity in cancer — using tissue repair as a road map. *Nat. Rev. Cancer*, **19**, 82–96.
32. Liu,C., Kanekiyo,T., Xu,H. and Bu,G. (2013) Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. *Nat. Rev. Neurol.*, **9**, 106–118.
33. Buchumenski,I., Roth,S.H., Kopel,E., Katsman,E., Feiglin,A., Levanon,E.Y. and Eisenberg,E. (2021) Global quantification exposes abundant low-level off-target activity by base editors. *Genome Res.*, **31**, 2354–2361.
34. Kannan,S., Altae-Tran,H., Jin,X., Madigan,V.J., Oshiro,R., Makarova,K.S., Koonin,E.V. and Zhang,F. (2021) Compact RNA editors with small Cas13 proteins. *Nature Biotechnology*, **40**, 194–197.
35. Xu,C., Zhou,Y., Xiao,Q., He,B., Geng,G., Wang,Z., Cao,B., Dong,X., Bai,W., Wang,Y. *et al.* (2021) Programmable RNA editing with compact CRISPR–Cas13 systems from uncultivated microbes. *Nat. Methods*, **18**, 499–506.
36. Katrekar,D., Xiang,Y., Palmer,N., Saha,A., Meluzzi,D. and Mali,P. (2022) Comprehensive interrogation of the ADAR2 deaminase domain for engineering enhanced RNA editing activity and specificity. *Elife*, **11**, e75555.
37. Han,W., Huang,W., Wei,T., Ye,Y., Mao,M. and Wang,Z. (2022) Programmable RNA base editing with a single gRNA-free enzyme. *Nucleic Acids Res.*, **50**, 9580–9595.
38. Rauch,S., He,E., Srienc,M., Zhou,H., Zhang,Z. and Dickinson,B.C. (2019) Programmable RNA-guided RNA effector proteins built from human parts.*Cell*, **178**, 122–134.

Supporting Information of Manuscript 1

Precise and efficient C-to-U RNA base editing with SNAP-CDAR-S

Ngadhnjim Latifi, Aline Maria Mack, Irem Tellioglu, Salvatore Di Giorgio and Thorsten Stafforst, *Nucleic Acids Research*, 2023, *51.15,* e84

Precise and efficient C-to-U RNA Base Editing with SNAP-CDAR-S

Ngadhnjim Latifi[1], Aline Maria Mack[1], Irem Tellioglu[2,3], Salvatore Di Giorgio[2] and Thorsten Stafforst[1,4*]

[1]Interfaculty Institute of Biochemistry, University of Tübingen, Auf der Morgenstelle 15, 72076 Tübingen, Germany
[2]Division of Immune Diversity (D150), German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany.
[3]Faculty of Engineering, University of Heidelberg, 69120 Heidelberg, Germany
[*]correspondence to thorsten.stafforst@uni-tuebingen.de
[4]Gene and RNA Therapy Center (GRTC), Faculty of Medicine University Tuebingen

Table of contents

Cells stably expressing editases

Generation of stable cell lines

For creation of cell lines stably expressing editases, the Flp-In™ T-REx™ system (Invitrogen, #R78007) was used. Hek 293 Flp-In™ T-Rex™ cells were cultivated in Dulbecco's modified eagle medium (DMEM; Life Technologies, #41965062) with fetal bovine serum (FBS; Life Technologies, #10270, 10% final conc.), Zeocin™ Selection Reagent (Z; Invitrogen, #R25001, 100 µg/ml final conc.), and Blasticidin S Hydrochloride (B; Blasticidin S Hydrochloride, Carl Roth, #CP14.2, 15 µg/ml final conc.) For generation of cell lines, 1.6 x 10^6 cells were seeded in a 6 cm dish in DMEM +10% FBS, +Z, +B. After 24 h medium was switched to DMEM + 10% FBS without antibiotics 1 h prior to transfection. Cells were transfected with 4 µg of plasmid in a 1:9 ratio of pOG44 Flp-Recombinase expression vector (invitrogen, #V600520) and pcDNA™5/FRT Mammalian Expression Vector (Invitrogen, #V601020) the corresponding editase with 12 µl Lipofectamine™ 2000 (ThermoFisher Scientific, #11668019) in 600µl Opti-minimal essential medium I (Opti-MEM™ I; Life Technologies, #11058021) final volume. After 24 h, medium was changed to selection medium swapping Zeocin with CELLPURE® Hygromycin B solution (H; Carl Roth, #CP12.1, 100 µg/ml final conc.) After 2 weeks of selection, cells were taken into culture in DMEM +10% FBS, +B, +H.

Cloning of constructs

All construct for generation of cell lines stably expressing them were cloned into the pcDNA™5/FRT Mammalian Expression Vector (Invitrogen, #V601020).

Apo1S constructs

We added localization tags to Apo1S at the C-terminus. For this, tags were ordered as synthetic oligonucleotides, which were annealed and ligated into the pcDNA™5. Oligo pairs (1 µl of either oligo 100 µM in H2O) were diluted in 200 µl of water. 39 µl of dilution were

SNAP-CDAR-S constructs

Mutations for transversion of ADARs adenine deamination activity to cytidine deamination (Abudayyeh, O. O., et al., Science 365.6451 (2019): 382-86) were transferred to the deaminase domain of our SNAP-ADAR2Q construct and synthesized by ThermoFisher Scientific GeneArt Services.

Microscopy

Glass cover slips were inserted in 24-well plate well and coated with Poly-(D)-lysine hydrobromide (Sigma-Aldrich, #P6407-5MG, 0.1 mg/ml final conc. in H_2O) for 30 min. After subsequent washing with H2O and Phosphate-buffered saline (PBS; 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4), the slips were dried for 30 min. under UV irradiation. After 30 min. more of drying, the cover slips were ready for use. 5 x 10^4 Flp-In™ T-Rex™ cells expressing SNAP-editases were seeded on the cover slips in DMEM, FBS (10%), B, H and Doxycycline hyclate BioChemica (D; PanReacAppliChem, #24390-14-5, 10 ng/ml final conc.) to induce expression. After 24 h, cells were washed with PBS and incubated for 30 min. with staining solution. Staining solution consisted of DMEM, FBS (10 %) containing 2 µl NucBlue™ Live ReadyProbes™ Reagent Hoechst33342 (ThermoFisher Scientific, #R37605) and acetylated benzylguanine fluorescein isothiocyanat (ac-BG-FITC, 2 µM final conc.) in 200 µl final volume. Cells were fixated by addition of 21.6 µl of 37 % of aqueous formaldehyde (3.7 % final conc.) and incubation for 3 min. Cells were then washed three times with PBS (3 x 200 µl) and permeabilized with 200 µl PBS containing Triton X-100 (Carl Roth, #3051.3, 0.1 % final conc.) by incubation for 15 min. For microscopy, cover slips were mounted on microscope slides using Dako Fluorescence Mounting Medium (AgilentDako, #S302380-2) and dried over-night at 4 °C. Images were taken with an AXIO Observer.Z1 (Zeiss), Colibri.2 light source under 63x magnification.

Editing with Snap-editases

snap/$(snap)_2$-gRNA synthesis

gRNAs were designed and purchased by Eurogentec or SigmaAldrich with a 5'-C6-Amino linker or a 3'-C7-Amino linker referred to as NH_2-gRNA. Pre-activation of snap-linker was carried out with EDCI. 120 nmol of snap-linker (2 µl of 60 mM in DMSO) were incubated with 224 nmol of EDCI (2 µl of 112 mM in DMSO), 306 nmol of NHS (2 µl of 153 mM in DMSO), and 2 µl of DIPEA (5% solution in DMSO) for 60 min at 30°C. After pre-activation, 4 µl of DIPEA (5% solution in DMSO) were added to 8.5 µl of NH2-gRNA solution (50 µg of 6 µg/µl in H2O). For conjugation, half of the pre-activation mix was added to the NH2-gRNA and incubated for 30 min at 30°C. The rest of the pre-activation mix was added and incubated for 60 min at 30°C.

Pre-activation of snap2-linker was carried out DIC. 120 nmol of snap2-linker (2 µl of 60 mM in DMSO) were incubated with 540 nmol of DIC (2 µl of 270 mM in DMSO), 920 nmol of NHS (2 µl of 460 mM in DMSO), and 2 µl of DIPEA (5% solution in DMSO) for 4 h at 45 °C or over-night at 37 °C. Mix was lyophilized to remove residual DIC and subsequently dissolved in 12 µl of DIPEA (1.7 % in DMSO). 8.5 µl of NH_2-gRNA solution (50 µg of 6 µg/µl in H_2O) was added to the pre-activated linker and incubated for 2 h at 37 °C.

PAGE

Successful coupling was determined by separation on a 20% 5M urea PAGE. Gels were cast in large glass plates separated by 0.8 mm spacers. Mix for gel was made with 204 ml of ROTHIPHORESE® Sequencing gel concentrate (25 %, Carl Roth, #3043.1), 13 ml of ROTHIPHORESE® Sequencing gel buffer concentrate (50 %, Carl Roth, #3050.1), 13 ml of H_2O, 650 µl of Ammonium peroxydisulphate (Carl Roth, #9592.3, 10% solution in H_20), and 65 µl of tetramethylehtylenediamine (TEMED; Carl Roth, #2367.3). Cast gels were allowed to solidify at room temperature over-night. Each sample was supplemented with 5 µl of RNA-loading dye, consisting of RITHOPHORESE® Sequencing gel diluent (50 %, Carl Roth, #3047.1) diluted 1:10 in H_2O supplemented with bromphenol blue sodium salt (Carl Roth, #A512.1) and xylene cyanole (Carl Roth, #A513.1). Gel was run in Tris-Boric Acid-EDTA buffer (TBE, Tris 8.9 mM, Boric Acid 8.9 mM, EDTA 0.2 mM) for about 6 h, at 65 W, 90 mA.

For visualization of bands, the gel was placed onto UV-reflecting TLC Silica gel 60 F_{254} plates (Sigma-Aldrich, #1055540001) wrapped in transparent plastic wrap and irradiated at 254 nm. guideRNAs absorbed the radiation and appeared as dark bands. guideRNAs bearing the snap/$(snap)_2$-linker run slower on the gel. As a reference, uncoupled NH_2-guideRNA was loaded in a separate well. Bands of correct size were excised, and gel slices were shaken over-night at 4°C in nuclease-free H_2O. Purification was done by ethanol precipitation. For this, 0.1 volumes of sodium acetate (3 M in H_2O) and 3.5 volumes of ethanol were added to the snap-gRNA solution and precipitated over-night at -20°C. Next, the mix was centrifuged for 60 min at 4°C with 14,000 RPM. Supernatant was discarded and pellet was washed with 500 µl of pre-chilled 70% Ethanol (in H_2O) and centrifuged again for 60 min at 4°C with 14,000 RPM. Pellet was dissolved in nuclease-free H_2O and concentration and purity were determined by NanoDrop™ 2000/2000c Spectrophotometers (ThermoFisher Scientific).

Transfection of Snap-editases expressing cells

3 x 10^5 HEK 293 Flp-In™ T-Rex™ cells stably expressing editases were seeded in a 24-well format in DMEM with 10% FBS supplemented with doxycycline (10 ng/ml final concentration) to induce transgene expression. After 24 h, cells were transfected with 300 ng of pcDNA 3.1 expressing transcript of interest with 0.9µl Lipofectamine™ 2000. 24 h post-transfected 8 x 10^4 cells were reverse transfected with 5 pmol (unless differently stated) of snap/$(snap)_2$-guideRNA and 0.75 µl Lipofectamine™ 2000 by pipetting the transfection mix in a 96-well format and dripping cell suspension onto it. Unless differently stated, editing endpoint was 48 h post guideRNA transfection. When endogenous targets were edited, the forward transfection of pDNA was omitted and gRNAs were transfected 24 h after seeding.

Editing with RESCUE-S

Cloning of guideRNA oligos

Oligos coding for the guideRNAs bearing the corresponding overhangs were cloned as described before (Abudayyeh, O. O., et al., Science 365.6451 (2019): 382-86). Multiple oligos were designed

based on reported designs or best reported designs were purchased. Oligos were annealed and phosphorylated. For this, 1 pmol of either oligo (1 µl of 100 µM in H_2O) were diluted in H_2O (200 µl final volume). 39 µl of dilution was incubated with 5 µl of ATP (10 mM, NewEngland Biolabs, #P0765S) and 10 units of T4 Polynucleotide Kinase (PNK; NewEngland Biolabs, #M0201L) in 50 µl final volume for 30 minutes at 37 °C. After that, mix was heated up to 95 °C and slowly cooled down to room-temperature to ensure correct annealing of phosphorylated oligos. 1 µg of pC0041 pDNA for gRNA expression (Addgene, #103852) was digested with *Bbs* I (NewEngland Biolabs) in 50 µl total volume for 1 h at 37 °C. Successful digest was determined by gel electrophoresis in a 1.4 % Agarose gel (ROTI®Garose Agarose NEEO ultra-quality, Carl Roth, #2267.4) in Tris-Acetate-EDTA-Buffer (TAE; 40 mM Tris, 20 mM Acetic Acid, 1 mM EDTA,) for 30 min. at 120 V. Correct bands were excised an isolated using the NucleoSpin® Gel and PCR Clean-up Mini kit (Macherey-Nagel, #740609.50) following manufacturer's instructions. 1 µl of oligo phosphorylation and annealing reaction mix was ligated into 30 ng of linearized vector using 0.5 µl of T4 DNA Ligase (NewEngland Biolabs, #M0202L) in 10 µl final volume by incubation at room temperature for 15 min. All of the mix was used for heat shock transformation of XL-1 Blue chemically competent bacteria. 50 µl of chemically competent bacteria were diluted in 100 µl of Tris-EDTA buffer (TE; 10 mM Tris-HCl, 1 mM Na-EDTA). All of the mix was added to the mix and incubated on ice for 30 min. After that, mix was incubated at 42 °C for 1 min. and further incubated on ice for 5 min. Bacteria were reactivated in 1 ml of LB-medium (25 g of LB Broth, Carl Roth, #X968.2 in 1 l of H_2O) and shaking at 37 °C for 1 h. Cells were then streaked on an Ampicillin sodium salt CELLPURE® (Amp; Carl Roth, #K029.1, 100 µg/ml final conc.) containing LB-medium plate and incubated at 37 °C over-night. Single colonies were picked and incubated in liquid LB-medium while shaking at 37 °C over-night. Plasmids were isolated from the liquid cultures by using the NucleoSpin® Plasmid Transfection-grade kit (MachereyNagel, #740490.250) following manufacturer's instructions.

Transfection of RESCUE-S cell lines

2 x 10^4 HEK 293 Flp-In™ T-Rex™ cells stably expressing RESCUE-S were seeded in a 96-well format in DMEM with 10% FBS supplemented with doxycycline (10 ng/ml final concentration) to induce

transgene expression. After 24 h cells were transfected, as previously described (Abudayyeh, O. O., et al., Science 365.6451 (2019): 382-86). Cells were transfected with 300 ng of guideRNA expressing pDNA and 40 ng of target/reporter pDNA with 0.5 µl Lipofectamine™ 2000. Editing endpoint was 48 h post guideRNA transfection. When endogenous targets were edited, target pDNA was excluded.

Analysis of editing yield

At endpoint, cells were lysed with RLT-buffer (Qiagen, #79216) and the total RNA was isolated using the Monarch® RNA Cleanup Kit 10 µg (New England BioLabs, #T2030L) following manufacturer's instructions. For DNA depletion, 1 µg of total RNA was incubated for 30 min at 37°C with DNase I (NewEngland BioLabs, #M0303L) in 25 µl total volume. Reaction was terminated by addition of 2 µl EDTA (25 mM) and incubation for 10 min. at 75 °C. 7.5 µl of reaction (~250 ng) was utilized for target site amplification using either the One Step RT-PCR Kit (BiotechRabbit, #BR0400102) or One*Taq*® One-Step RT-PCR Kit (NewEngland BioLabs, #E5315S) with the according primers in 25 µl total volume. Correct amplicon size was determined by agarose gel electrophoresis. For this, samples were loaded on a 1.4 % agarose gel in TAE (> 1,000 bp) or Sodium-Borate-Buffer (SB; 10 mM NaOH, 36.5 mM Boric Acid, <1,000 bp amplicon) and separated at 120 V for 30 min (TAE) or 200 V for 15 min (SB). Correct amplicons were excised and isolated using the NucleoSpin® Gel and PCR Clean-up Mini kit following manufacturer's instructions. 100ng of amplicon was sequenced with the respective primer. Sanger Sequencing was carried out by Microsynth or Eurofins.

Editing analysis of APOE

APOE is a high GC-content transcript. Therefore, downstream preparation had to be adjusted. Post RNA isolation, 1 µg of total RNA was DNA-depleted with TURBO DNA-free™ kit (ThermoFisher Scientific, #AM1907) in 30 µl total volume (1 µl of TURBO™ DNase) for 30 min. at 37 °C. Reaction was terminated by addition of 2 µl of Turbo DNase™ Inactivation buffer and incubation at room-temperature for 5 min. 5 µl of mix was then used for RT conducted using the

SuperScript™ IV First-Strand Synthesis System (ThermoFisher Scientific, #18091050). Mix was incubated with 1 µl of dNTPs, 2 µl of random primer mix (10x), and 1 µl of gRNA sense strand oligo (10 µM) at 95 °C for 5 min. and immediately afterwards cooled down on ice. Next, RT was set up by addition of 4 µl of SuperScript™ IV buffer (5x), 1 µl DTT (0.1 M), 1 µl RNase Inhibitor (murine), and SuperScript™ IV RT in 20 µl final volume. 5 µl of reaction mix was used for PCR with Taq Polymerase (NEB, #M0267S) containing DMSO (10% final conc.) in 50 µl final volume. Both, determination of correct amplicon size and its excision was conducted as described above. Sequencing was also performed as described above.

CTNNB1 Assay

Transfection conditions

3 x 10^5 HEK 293 FlpIn TRex cells stably expressing SNAP-CDAR-S were seeded in a 24-well format in DMEM with 10% FBS supplemented with doxycycline (10 ng/ml final concentration) to induce transgene expression. After 24 h cells were forward transfected with 300 ng of pcDNA 3.1 expressing Renilla Luciferase and 300 ng of either TOPFlash (FireFly Luciferase expression with TCF/LEF responsive promoter elements; Addgene #12456) or FOPFlash (FireFly Luciferase expression with mutated TCF/LEF responsive promoter elements; Addgene #12457) pDNA with 0.9 µl (per 300 ng pDNA) of Lipofectamine™ 2000 (Life Technologies). 24 h after pDNA transfection, 8 x 10^4 cells were either reverse transfected as described with 5 pmol of either CTNNB1 T41-targeting NH_2-gRNA, CTNNB1 T41-targeting $(snap)_2$-gRNA, PPIB R7C-targeting $(snap)_2$-gRNA or without any gRNA in technical duplicates. 48 h after gRNA transfection FireFly expression was measured with the Dual-Luciferase® Reporter Assay System by Promega according to manufacturer's instructions. Cells were lysed with 30 µl of Passive Lysis buffer per well (96-well format) and shaken for 15 min. at room temperature.

For RESCUE-S transfection, 2 x 10^4 HEK 293 Flp-In™ T-Rex™ cells stably expressing RESCUE-S were seeded in a 96-well format in DMEM with 10% FBS supplemented with doxycycline (10 ng/ml final concentration) to induce transgene expression. After 24 h cells were transfected, as previously described (Abudayyeh, O. O., et al., Science 365.6451 (2019): 382-86). Briefly, cells

were transfected with 300 ng of pDNA expressing either CTNNB1 T41-targeting, PPIB R7C-targeting or no gRNA at all, 20 ng of pcDNA3.1 expressing Renilla Luciferase and 20 ng of TOPFlash or FOPFLash pDNA with 0.5 µl Lipofectamine™ 2000 (Life Technologies) in technical quadruplicates. Another sample was included transfected only with Luciferase expressing pDNA. 48 h after gRNA transfection, FireFly expression was measured with the Dual-Luciferase® Reporter Assay System (Promega, E1910) according to manufacturer's instructions. Cells were lysed with 20 µl (per well) of Passive Lysis buffer per well (96-well format) and shaken for 15 min. at room temperature and two technical replicates were pooled.

FireFly Luciferase expression and editing yield

10 µl of each replicate was measured in a LumiNunc 96-well plate (VWR) with a Spark 10 M plate reader (Tecan). 35 µl per well of each substrate was added by an auto-injector in sequence. For FireFly Luciferase signal measurement, 35 µl per well of Luciferase Assay Reagent II (LAR II, Promega) were injected and incubated for 5 seconds, after which signal was measured for 10 seconds. For Renilla Luciferase signal measurement and FireFly Luciferase signal quenching, 35 µl per well of Stop & Glo® Reagent and incubated for 5 seconds, after which signal was measured for 10 seconds.

Technical duplicates were pooled, and total RNA was isolated using the Monarch® RNA Cleanup Kit 10 µg (New England BioLabs) following manufacturer's instructions. Editing yield was determined as described above.

FireFly signal and editing yield were determined in three biological replicates.

STAT3 Assay

3 x 10^5 HEK 293 FlpIn TRex cells stably expressing editases were seeded in a 24-well format in DMEM with 10% FBS supplemented with doxycycline (10 ng/ml final concentration) to induce transgene expression. After 24h, 3.2 x 10^5 cells were reverse transfected with 20 pmol of STAT3 S727F-targeting NH_2-gRNA, $(snap)_2$-gRNA, or PPIB R7C-targeting $(snap)_2$-gRNA (quadruple of 96-

well format) using 2 µl Lipofectamine™ RNAiMAX (invitrogen) per transfection. Transfection was repeated in the same fashion 2, 4, and 6 days after the first transfection. Endpoint was at day 8 after the first transfection. Cells were harvested and 20% were used for RNA isolation and editing analysis as described above. For cell lysis 10 ml of RIPA lysis and extraction buffer (ThermoFisher Scientific) was supplemented with one tablet of c0mplete™ Tablets, Mini, EDTA-free *EASY*pack Protease Inhibitor Cocktail (Roche) and PhosStop *EASY*pack (Roche), respectively. The rest of the cells was lysed in 50 µl of lysis buffer, incubated on ice for 15 min, and centrifuged for 10 min. at 4 °C and 14,000 RPM. Supernatant was collected and protein concentration was measured by Pierce™ BCA Protein Assay Kit (ThermFisher Scientific) in a Tecan Plate Reader (gucken wie das eigentlich heißt). 30 µg of protein were loaded in duplicates on a Novex™ WedgeWell™ 8 to 16%, Tris-Glycine, 1.0 mm, Mini Protein Gel (ThermoFisher Scientific) and separated for 60 min. at 200 V in 1:10 ROTHIPHORESE® 10X SDS-PAGE (Carl Roth, #3060):H2O. Blotting was performed with the Mini Trans-Blot Cell® (BioRad) in TransferBuffer (190 mM glycine, 25 mM Tris, 20% Methanol) for 60 min. at 100 V. Blocking was performed for 1 h at room temperature in blocking buffer consisting of Tris-buffered saline (TBS, 50 mM Tris, 150 mM NaCl) with Tween® 20 (VWR, #M147-1L, 1 % final conc.; TBST) containing Milk, non fat (skimmed milk), powder (VWR, 5 % final concentration). All antibodies were diluted in blocking buffer. The blot was cut in half and one side was used for total STAT3 detection and the other side was used for STAT3 pS727 detection. The former was incubated with Stat3 (DRZ2G) Rabbit mAb (CellSignaling) in a 1:1000 dilution and Monoclonal Anti-β-Actin antibody produced in mouse (Sigma) in a 1:5,000 dilution. The latter was incubated with P-Stat3 (S727) (D8C2Z) Rabbit mAb (CellSignaling) in a 1:1,000 dilution and Monoclonal Anti-β-Actin antibody produced in mouse (Sigma) in a 1:5,000 dilution. Incubation was performed for 3 days at 4 °C. Blots were washed three times for 5 min. with TBST. Secondary antibody incubation was performed using Peroxidase AffiniPure Goat Anti-Mouse IgG (Jackson Immuno Research, #115-035-003), and Peroxidase AffiniPure Goat Anti-Rabbit IgG (Jackson Immuno Research, #111-035-003) dissolved in blocking buffer at a 1:10,000 dilution for 90 min. at room temperature. Blots were washed again with TBST three times for 5 min. Blots were incubated with detection solution (100 mM Tris-HCl, 0.022 % Luminol, 0.0033 % p-coumaric acid) Images were taken with Odyssey FC Imager (Li-Cor® Biosciences).

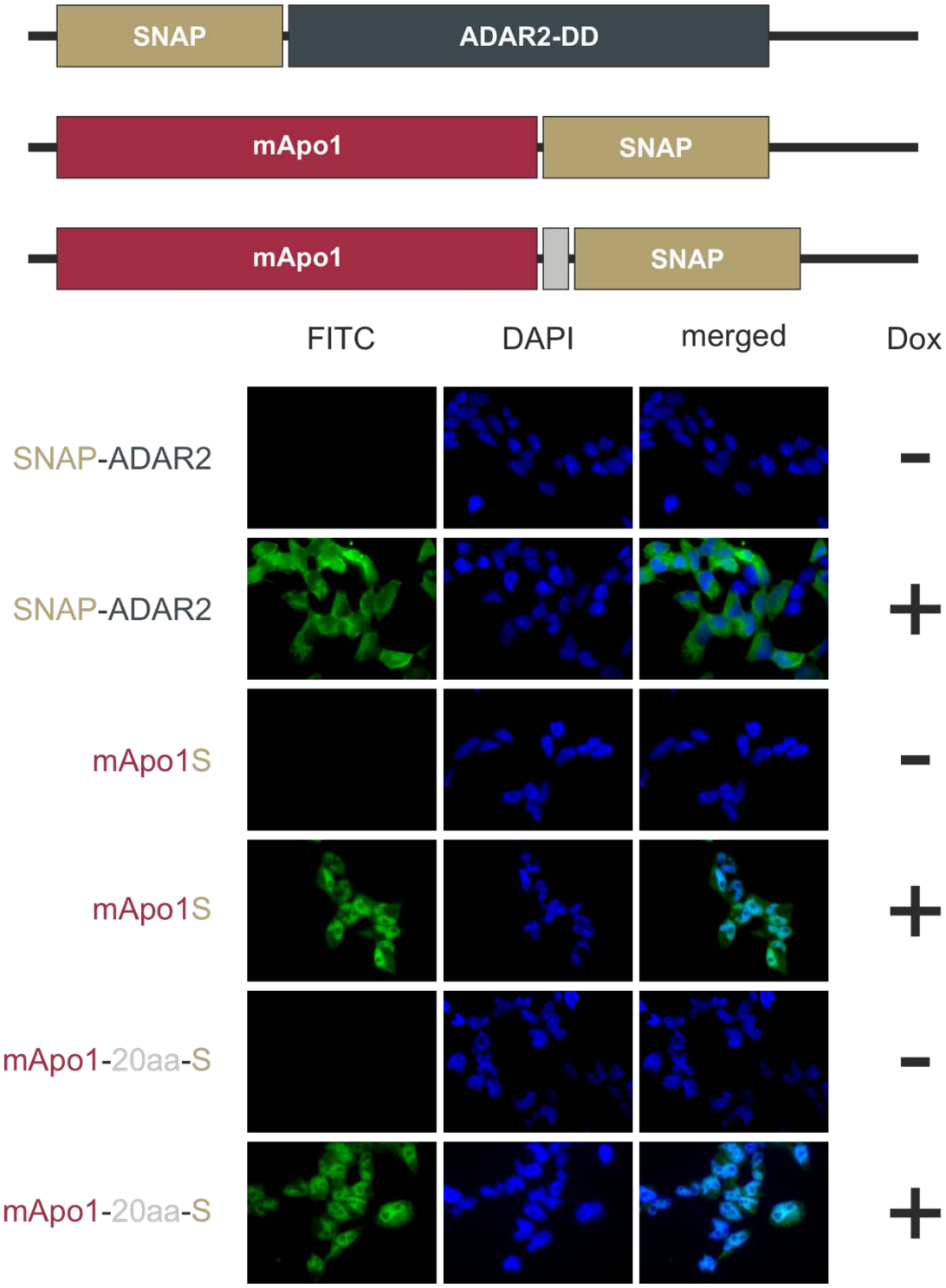
Next Generation Sequencing

For Next Generation Sequencing (NGS), samples transfected without anything or with PPIB R7C-targeting gRNA (2.5 pmol)/gRNA expression vector (300 ng) in technical duplicates. For this, cells expressing editases were transfected in a five-fold set up, as described above. 48 h after transfection, cells were lysed with 50 µl per well of RLT buffer (Qiagen) and corresponding wells were pooled (250 µl total volume). 200 µl of RLT buffer was added and the RNA was isolated using the RNeasy MinElute Cleanup Kit (Qiagen, #74204) following manufacturer's instructions for large scale. RNA was eluted in 30 µl of nuclease-free H_2O and concentration was determined by Nanodrop. Sasmples were DNA-depleted by DNase I (NewEngland BioLabs) digestion. For this, entire RNA was incubated with DNase I (1 µl per 2 µg of total RNA) for 30 min. at 37 °C. RNA was then purified using the RNeasy MinElute Cleanup Kit following manufacturer's instructions for small scale. RNA was then eluted in 30 µl of nucleas-free H_2O. Concentration was adjusted to 150 ng/µl and determined by Nanodrop in duplicate. Successful execution of editing was determined by Sanger Sequencing. 500 ng of total RNA was used for amplification of target site by One*Taq*® One-Step RT-PCR Kit (NewEngland BioLabs) following manufacturer's instructions. Sample preparation for sequencing was performed as described above. The samples were submitted in their entirety for NGS performed by CeGaT. Library preparation was conducted with 100 ng of RNA with the TruSeq Stranded mRNA (Illumina). Sequencing was performed 2 x 100 bp pair-end with NovaSeq6000 (Illumina) with 50 Mio reads. Lanes of raw RNA sequencing data of same samples were pulled together, and adapter sequences were trimmed with Trim Galore (v0.6.5; http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Sequencing alignment to the human reference genome (hg19) was performed using STAR (v. 2.7.10a). hg19 and the RefSeq annotation are publicly available at the genome browser at UCSC. For alignment uniquely mapped (STAR option: --outFilterMultimapNmax 1) reads were considered to prevent multimapping of regions of high similarity. Next, data (bam files) were deduplicated, sorted, and indexed using SAMtools (v1.9; http://samtools.sourceforge.net). SNVs calling was performed with REDItools (v2; https://github.com/tflati/reditools2.0). Developers' recommendations were taken into consideration for preceding data preparation. As previously performed, only high-

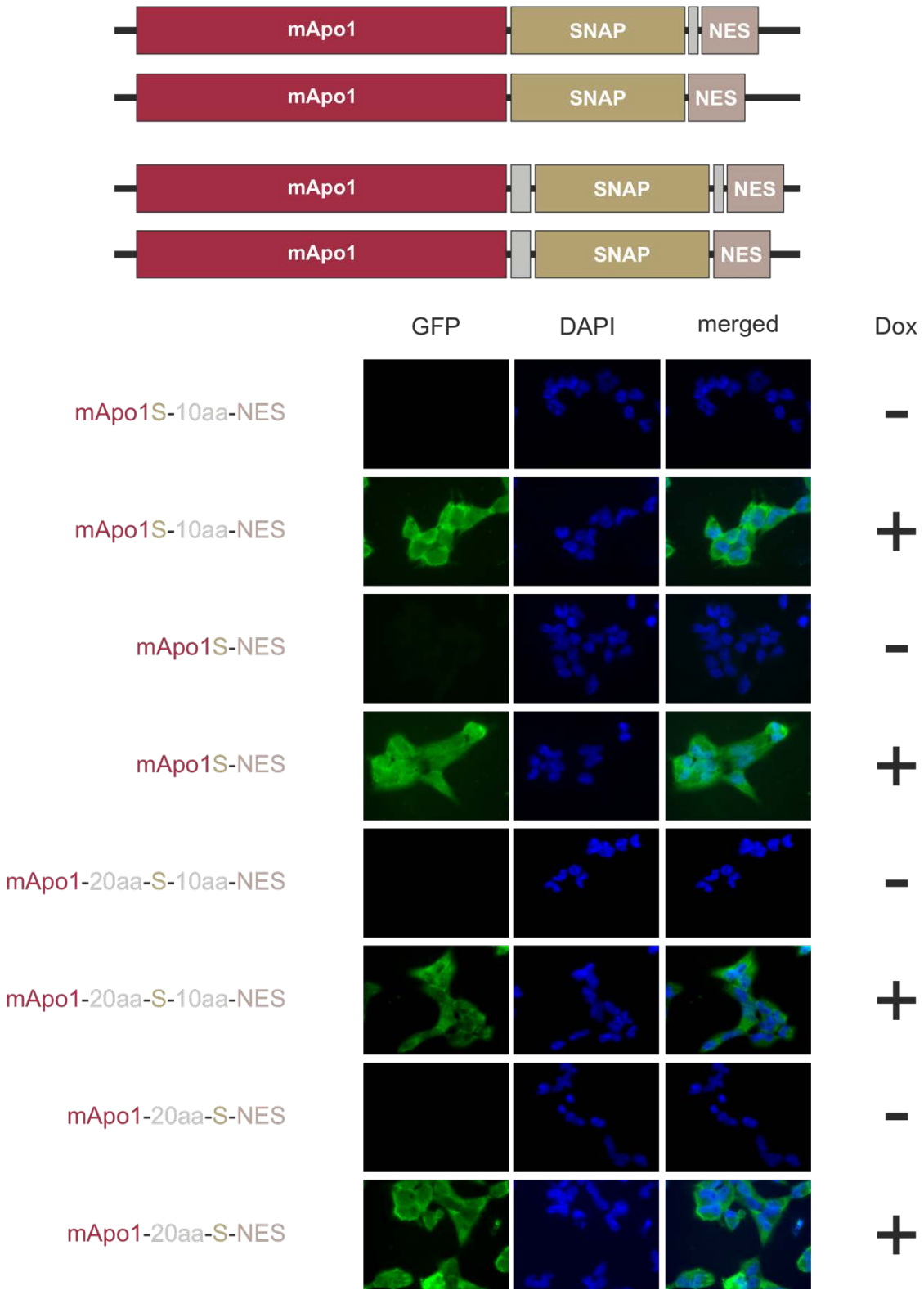
quality sites (min. MeanQ > 30 in REDItools2) were considered. Editing site were called when well-covered i.e., minimum 50 reads (in summary of duplicates per sample), showing ≥ 5 % editing frequency compared to the control. For sites matching criteria, fisher's exact test was performed. Significance was defined for all samples showing an adjusted *p*-value < 0.01. Genomic coordinates were annotated with Variant Effect Predictor (VEP) (v102) using the following command line:

```
vep –input file input..txt --fasta reference.fa --output_file output.txt --species homo_sapiens --tab --cache --dir_cache ../Human/dir --no_check_variants_order --transcript_version --canonical --ccds --hgvs --symbol --gene_phenotype --pubmed --variant_class --pick --offline --force_overwrite
```

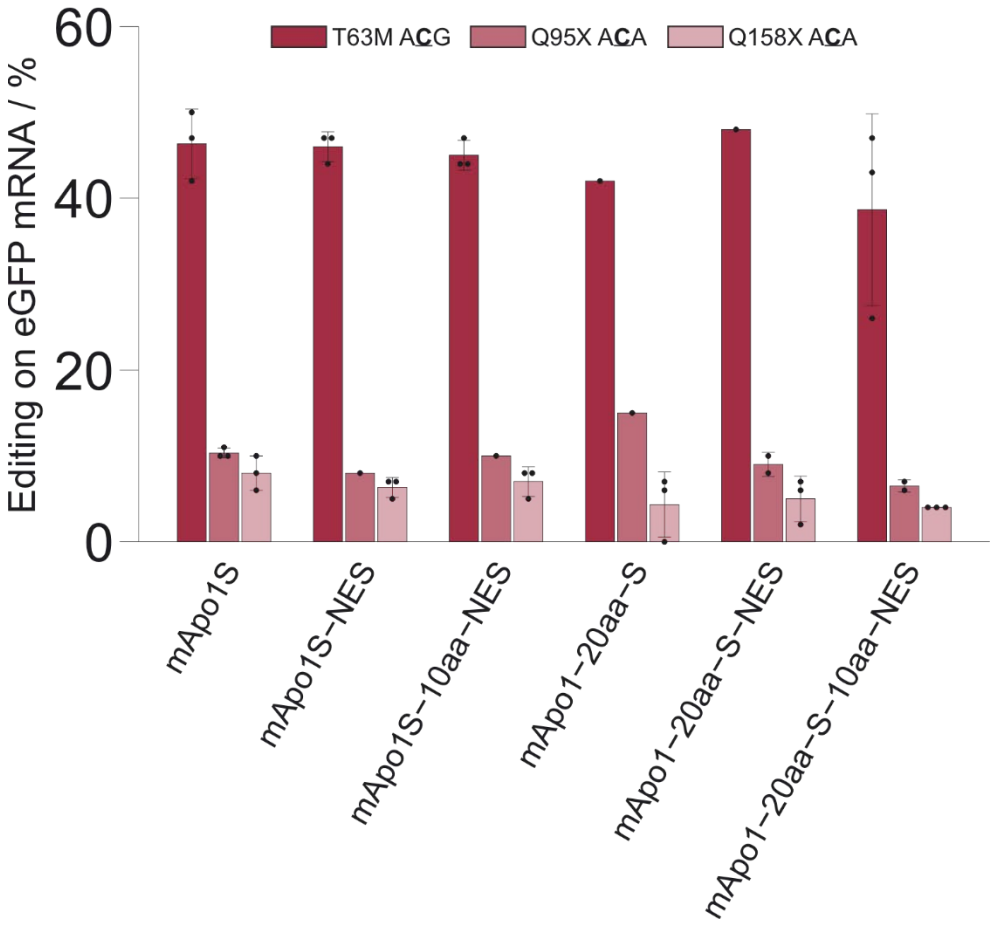
Supplementary Figures



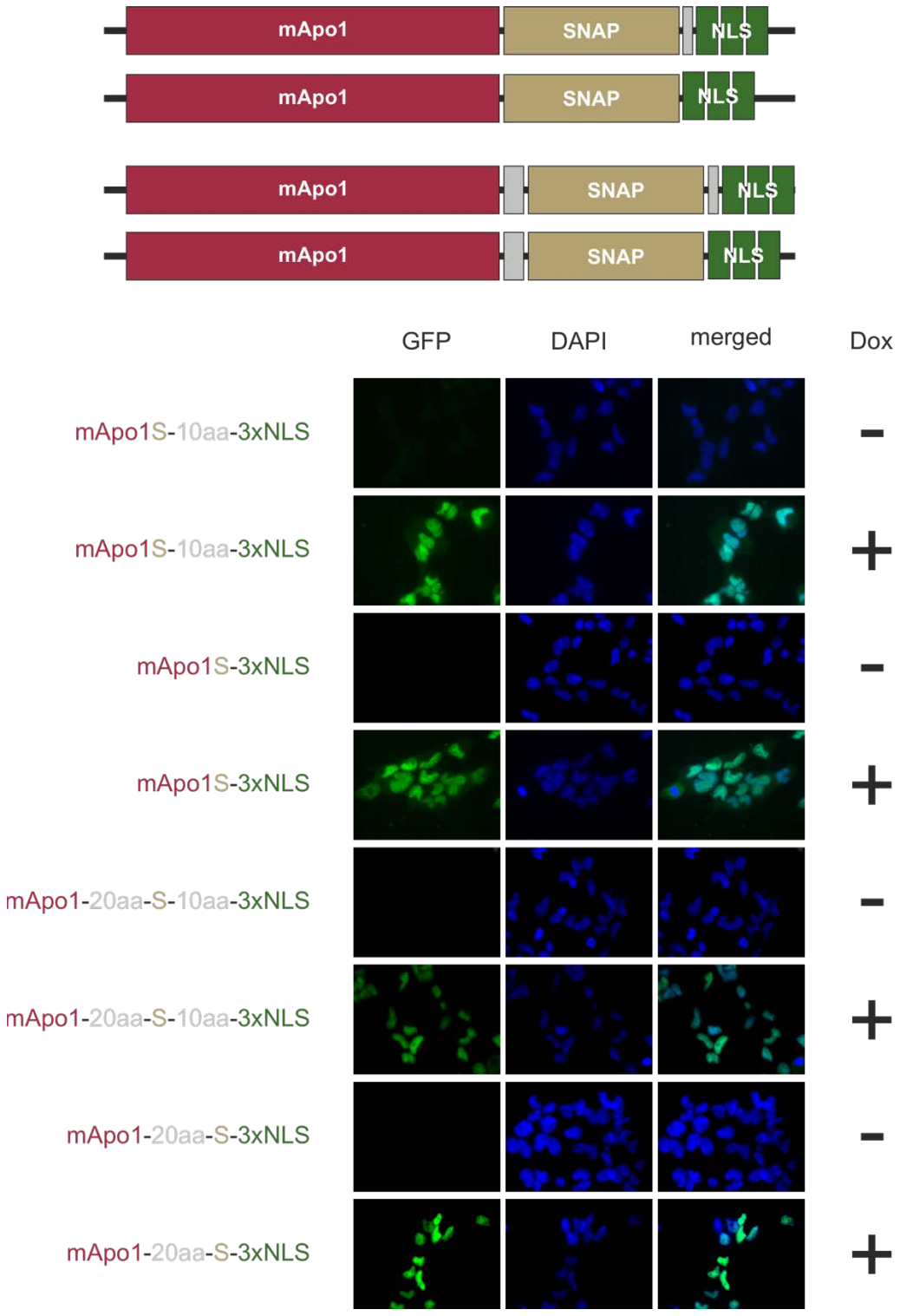
S. Figure S1: Microscopy of SNAP-tagged effectors. 5 x 10^4 cells were plated on glass cover slips and SNAP-effector expression was induced for 24 h with doxycycline. Images were taken at 24 h after doxycycline induction. SNAP-effectors were visualized by incubation with BG-FITC and DAPI stain was utilized for nucleus visualization. SNAP-ADAR2 served as control for cytoplasmatic localization.



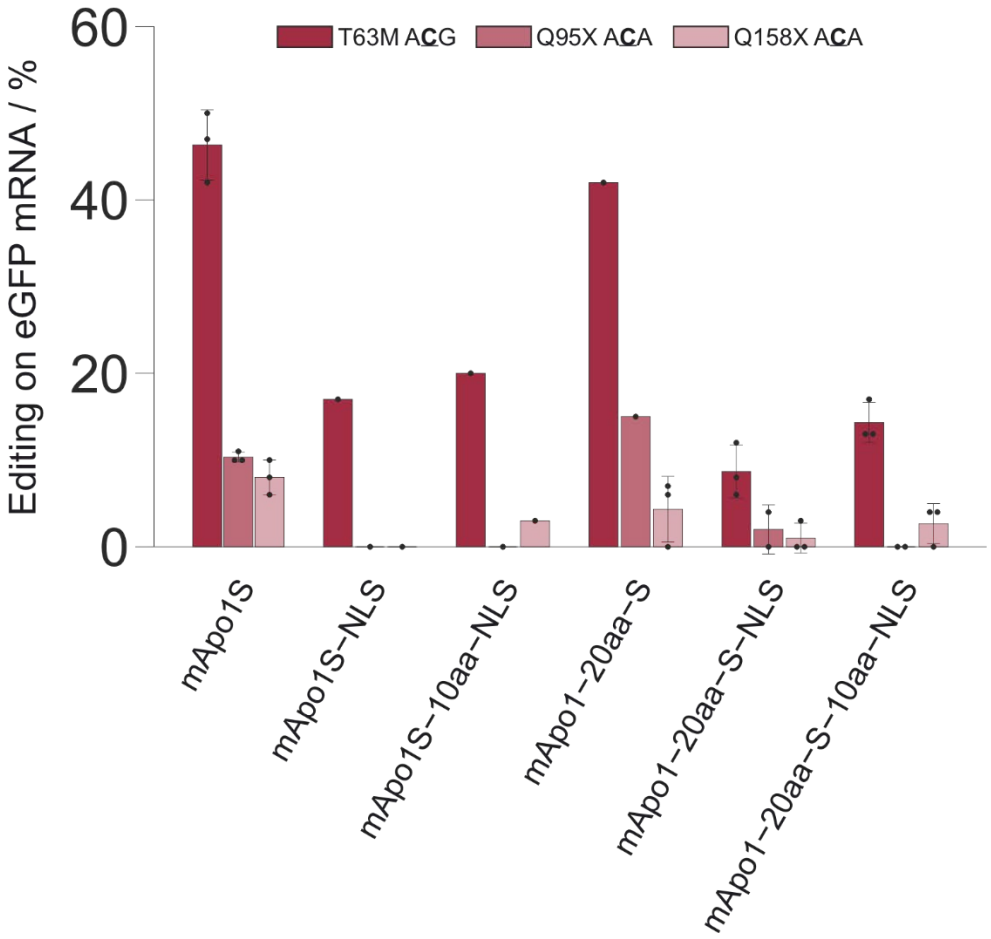
S. Figure S2: Microscopy of mApo1S constructs bearing a nuclear export signal (NES) of the HIV Rev protein. 5 x 10^4 cells were plated on glass cover slips and SNAP-effector expression was induced for 24 h with doxycycline. Images were taken at 24 h after doxycycline induction. SNAP-effectors were visualized by incubation with BG-FITC and DAPI stain was utilized for nucleus visualization.



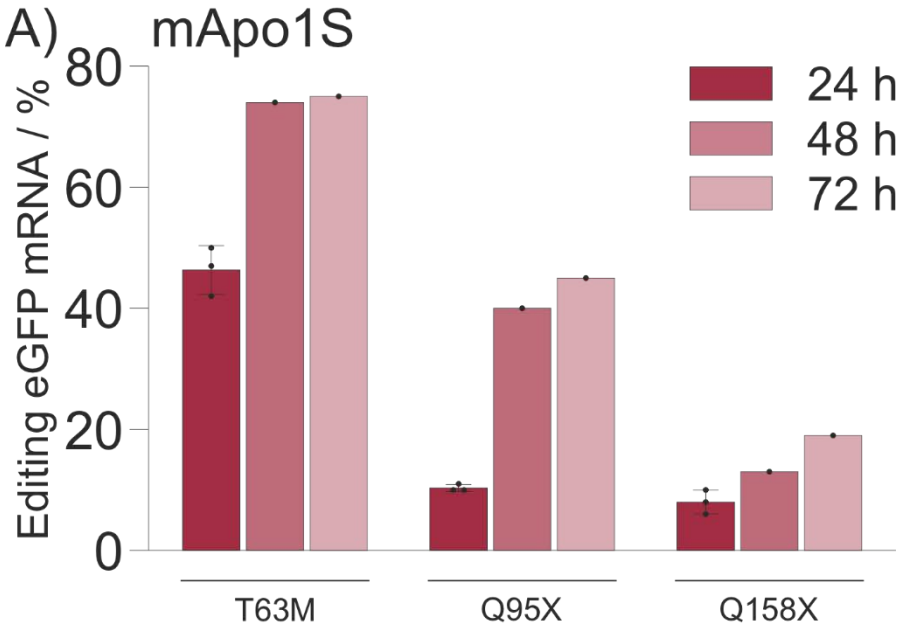
S. Figure S3: Effect of localization tags on Editing. 8 x 10^4 Cells stably expressing mApo1S constructs were transfected with 5 pmol of $(snap)_2$-gRNA targeting eGFP T63, Q95, or Q158, respectively. Editing yield was determined 24 h post-transfection. Data is shown as mean ± s.d. of N ≥ 1 of independent experiments, as indicated by the individual data points.

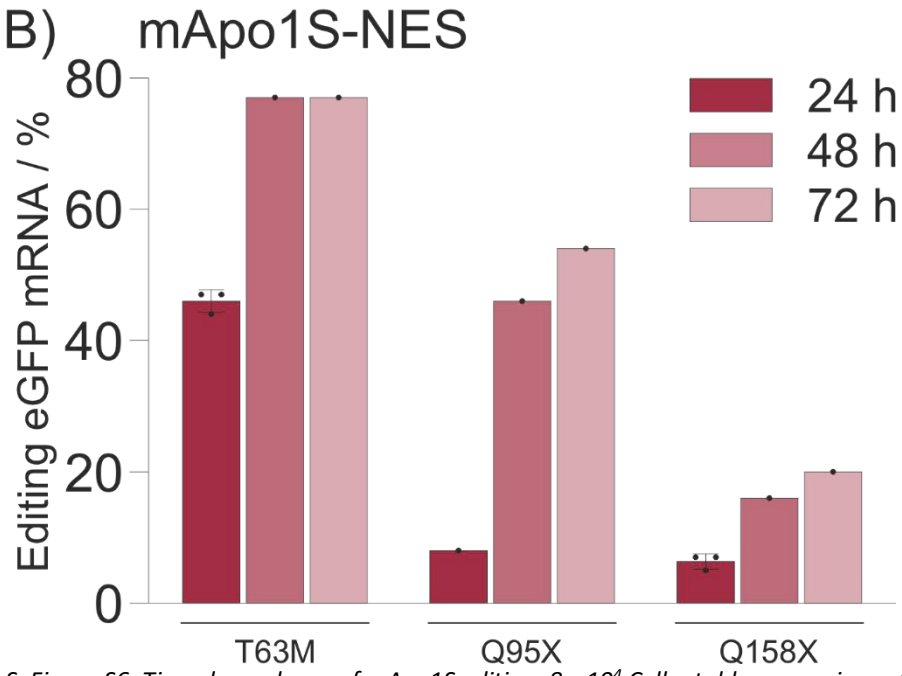


S. Figure S4: Microscopy of mApo1S constructs bearing three copies of nuclear localization signals of the SV40 large T antigen. Microscopy of mApo1S constructs bearing a nuclear export signal (NES) of the HIV Rev protein. 5 x 10^4 cells were plated on glass cover slips and SNAP-effector expression was induced for 24 h with doxycycline. Images were taken at 24 h after doxycycline induction. SNAP-effectors were visualized by incubation with BG-FITC and DAPI stain was utilized for nucleus visualization.



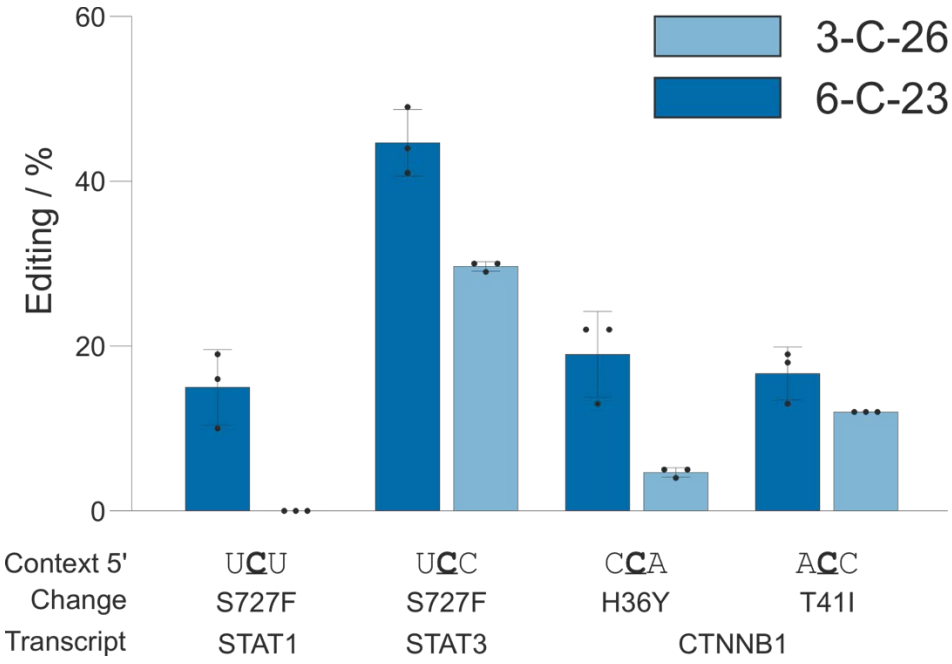
S. Figure S5: Effect of localization tags on Editing. 8 x 10^4 Cells stably expressing mApo1S constructs were transfected with 5 pmol of $(snap)_2$-gRNA targeting eGFP T63, Q95, or Q158, respectively. Editing yield was determined 24 h post-transfection. Data is shown as mean ± s.d. of N ≥ 1 of independent experiments as indicated by individual data points.



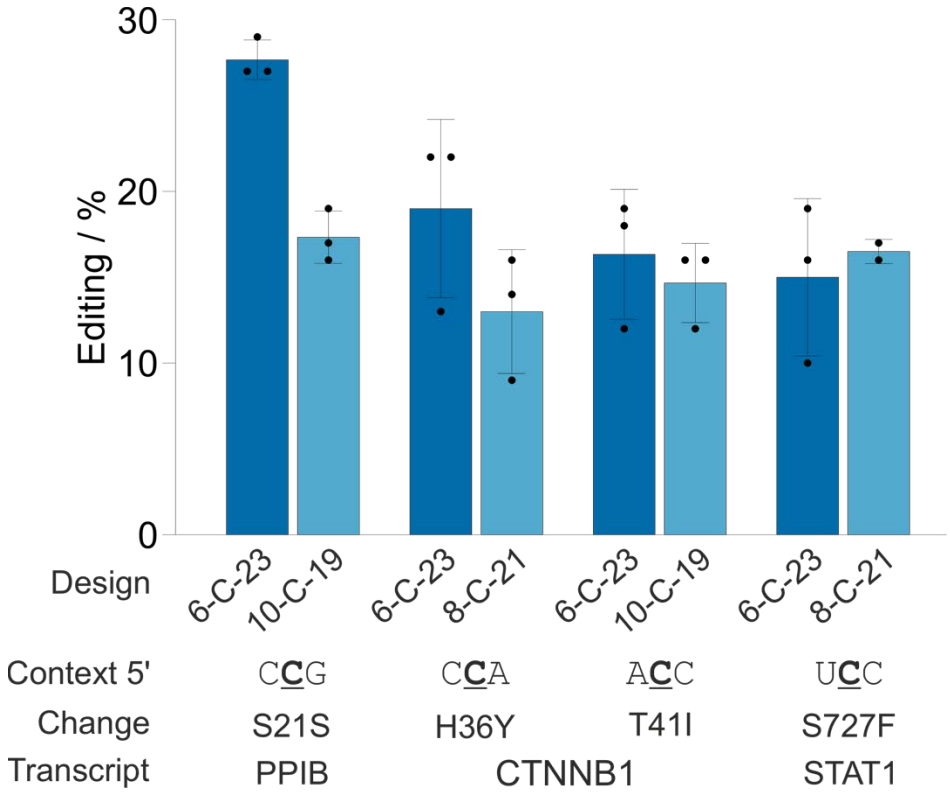


S. Figure S6: Time dependence of mApo1S editing. 8 x 10^4 Cells stably expressing mApo1S (A) or mApo1s-NES (B) were transfected with 5 pmol of $(snap)_2$-gRNA targeting eGFP T63, Q95, or Q158, respectively. Editing yield was determined at indicated time points post-transfection. Data is shown as mean ± s.d. of N ≥ 1 of independent experiments, as indicated by individual data points.

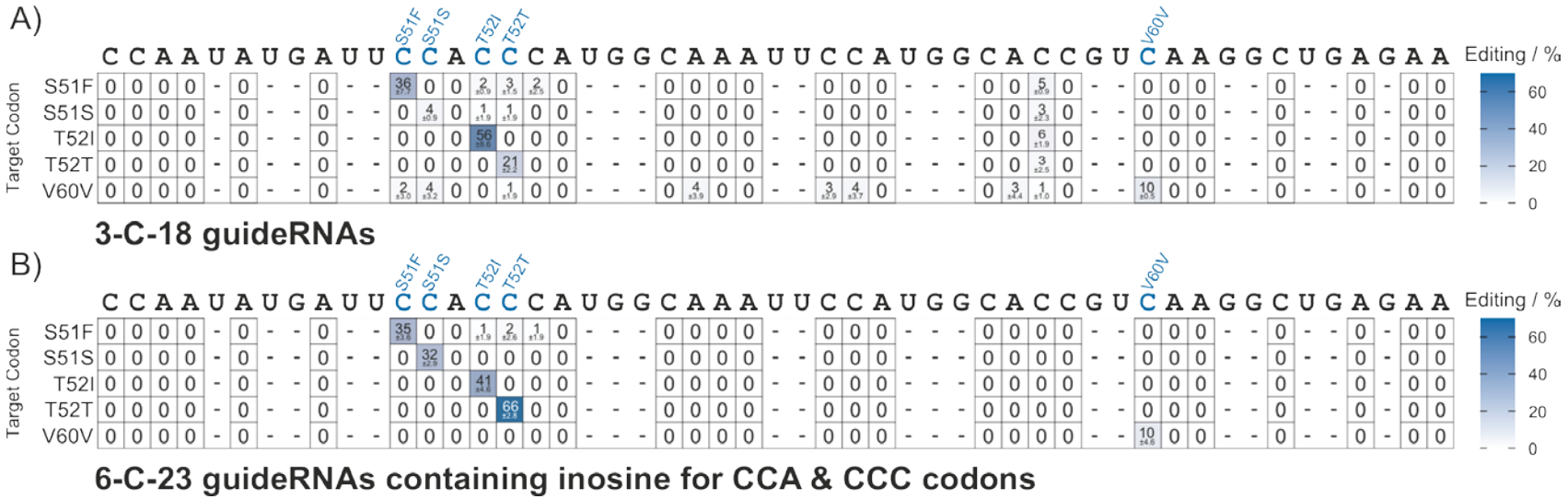
S. Figure S7: mApo1S editing of GAPDH T52. 8 x 10^4 cells were transfected with 5 pmol of $(snap)_2$-gRNA targeting GAPDH T52. Endpoint was 48 h post-transfection. On-target editing is shown in duplicates of independent experiments by red asterisk. Yellow asterisk indicates a persistent guideRNA-dependent by-stander editing, which almost exceeds the on-target editing yield and seems to be promoted by a local RNA stem-loop hairpin structure



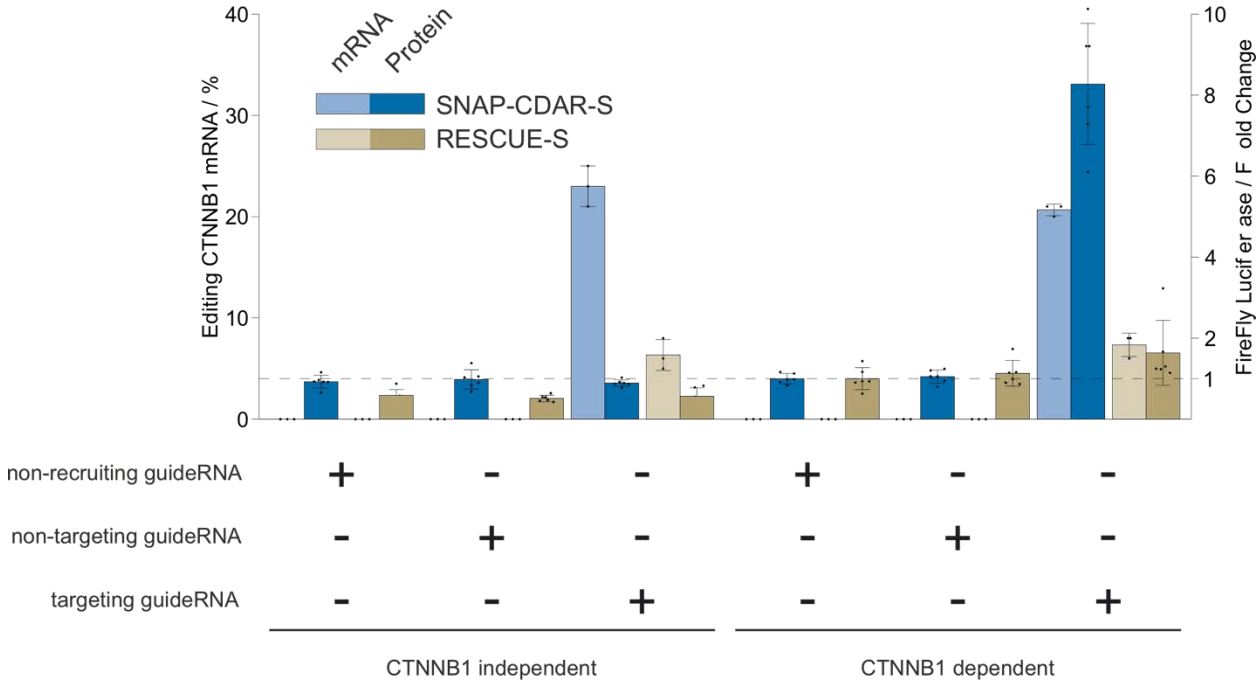
S. Figure S8: Comparison of gRNA designs. 8 x 10^4 cells stably expressing SNAP-CDAR-S were transfected with 5 pmol of $(snap)_2$-gRNA of indicated design, respectively. Best performing guideRNA design on PPIB (3-C-26) was compared to 6-C-23 design on various endogenous transcripts, indicating that the 6-C-23 design works typically better, with PPIB as the exception. Data is shown as mean ± s.d. of N = 3 of independent experiments.



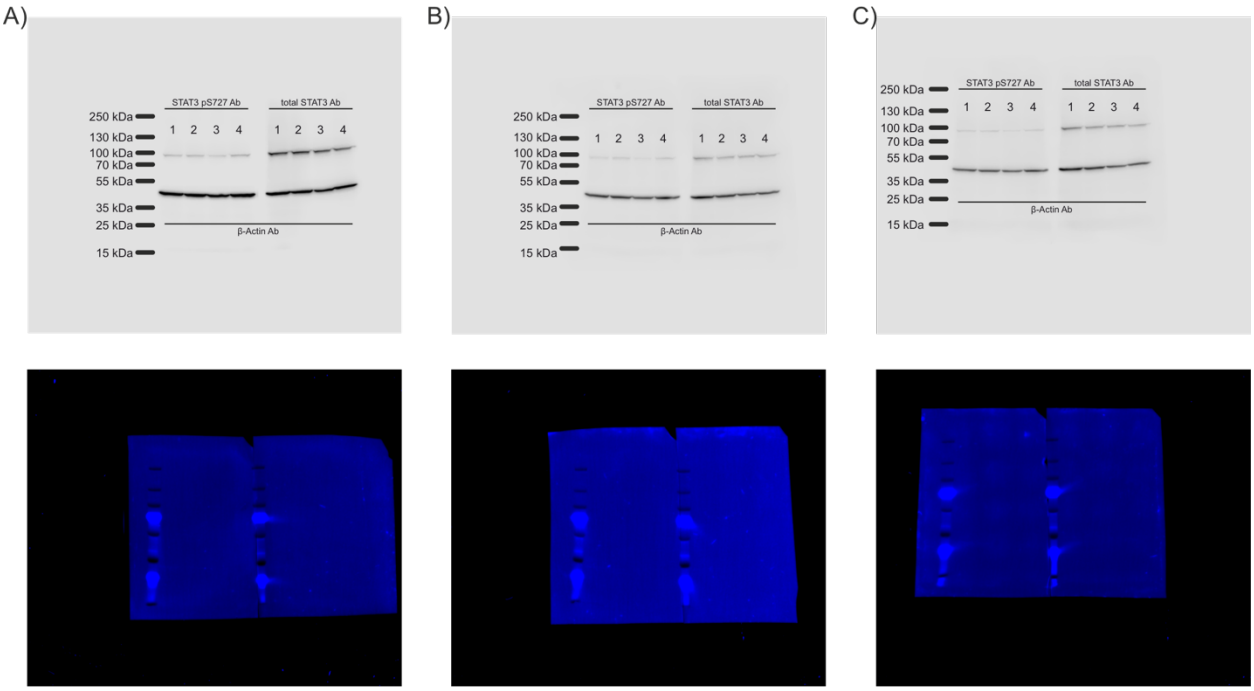
S. Figure S9: Comparison of lead gRNA designs with best reported design. 8 x 10^4 cells stably expressing SNAP-CDAR-S were transfected with 5 pmol of $(snap)_2$-gRNA of indicated design, respectively. SNAP-CDAR-S lead design (6-C-23) was compared to $(snap)_2$-gRNAs with designs reported best for RESCUE-S (position of mismatch C). Data is shown as mean ± s.d. of N = 3 of independent experiments.



S. Figure S10: Heatmap for Programmability of SNAP-CDAR-S. 8 x 10^4 cells stably expressing SNAP-CDAR-S were transfected with 5 pmol of either 3-C-18 (A) or 6-C-23 (B) $(snap)_2$-gRNAs targeting indicated target sites on GAPDH transcript. B) gRNAs targeting 5'CCN codons contained inosine opposite of 5'C. Depicted are a 51 bp area of GPADH transcript surrounding the target sites. Increase of on-target editing by the 6-C-23 design did not result in reduction of fidelity, as editing of by-stander sites remained at background level. Data is shown as mean ± s.d. of N = 3 of independent experiments.



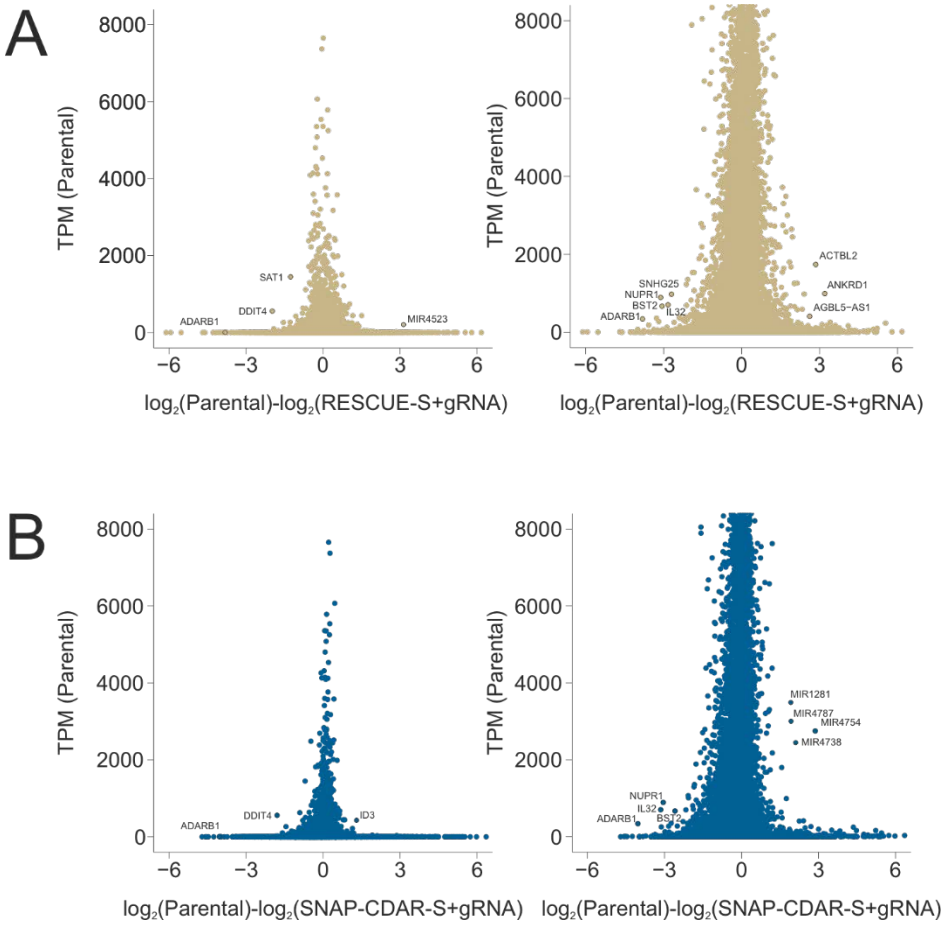
S. Figure S11: FireFly Luciferase Assay of CTNNB1 T41 Editing. FireFly Luciferase signal as fold change of cells transfected with a CTNNB1-dependent reporter plasmid DNA and a non-recruiting guideRNA. Definition of "non-recruiting guideRNA" was system-dependent (i.e., CTNNB1 T41-targeting NH_2-guideRNA for SNAP-CDAR-S and empty guideRNA expression plasmid for RESCUE-S). Data is shown as mean ± s.d. of N = 3 independent experiments. FireFly Luciferase signal was measured in technical duplicates of independent experiments, respectively.



Legend

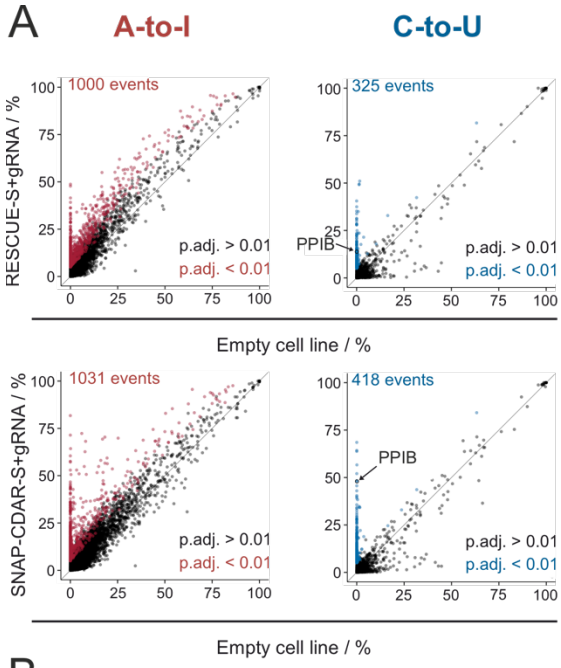
1: no guideRNA
2: STAT3 S727F NH_2-guideRNA
3: STAT3 S727F $(snap)_2$-guideRNA
4: PPIB R7C $(snap)_2$-guideRNA

S. Figure S12: STAT3 pS727 Editing. SNAP-CDAR-S stably expressing cells were transfected every 48 h with 5 pmol of (snap)2-gRNA targeting STAT3 S727 over a time period of 8 days. At endpoint, 30 µg of protein lysate was applied in duplicates. Lower row: Visualization of protein marker. Upper row: Visualization of phosphorylated S727 of STAT3 (left part of gels) and total STAT3 (right part of gels). The experiment was done in a triplicate of three independent experiments (A-C).



S. Figure S13: Volcano Plots of differential TPM-values of transcribed RNAs. A) RESCUE-S expressing cell lines. B) SNAP-CDAR-S expressing cell lines.

A



Empty cell line vs. Enzyme (+gRNA)	RESCUE-S		SNAP-CDAR-S	
	A-to-I	C-to-U	A-to-I	C-to-U
5'UTR	12	14	10	19
non-coding	201	26	203	53
non coding RNA	9	4	15	7
missense	192	82	178	83
stop lost	3	0	7	0
stop gained	0	16	0	13
silent	89	63	106	62
3'UTR	494	120	512	181
total	**1000**	**325**	**1031**	**418**

B

Table 1 | Global off-target editing

						Location in mRNA				
		Known		Novel			Coding region			
Enzyme[a]	Total	Alu	Non-Alu	Alu	Non-Alu	5′-UTR	Syn.	Nonsyn.	3′-UTR	Others[b]
SA1	6	2	1	0	3	0	0	1	3	2
SA2	30	15	8	1	6	0	0	2	22	6
SA1Q	835	70	59	7	699	11	117	230	402	75
SA2Q	1,310	267	71	24	948	13	149	347	637	164

Numbers represent the number of sites that were significantly differently edited compared with sites in a related control cell line that did not express the respective SA editase. Syn., synonymous; nonsyn., nonsynonymous. [a]Editing was carried out in cells expressing the given SNAP–ADAR in the presence of a BG-gRNA targeting the *ACTB* transcript. [b]"Others" refers to editing in introns, intergenic regions, and noncoding RNA.

S. Figure S14: Comparison of transcriptome-wide off-target sites. A) A-to-I and C-to-U transcriptome-wide off-target sites caused by SNAP-CDAR-S or Cas13-based RESCUE-S under stable integration of the respective effector into 293 Flp-In T-REx cell lines. B) For comparison, characterization of transcriptome-wide A-to-I off-target sites caused by stable integration of different SNAP-ADAR effectors in 293 Flp-In T-REx cell lines (SA1: wildtype SNAP-ADAR1; SA2: wildtype SNAP-ADAR2; SA1Q: hyperactive SNAP-ADAR1Q; SA2Q: hyperactive SNAP-ADAR2Q). The data indicates that the RESCUE-S domain, evolved form ADAR2Q, has still considerable A-to-I off-target effects, clearly above that of the wildtype ADAR2 deaminase. Table was taken from original publication by Paul Vogel et al. Nature Methods 2018.

5.4.2. Manuscript 2 (published)

Harnessing self-labeling enzymes for selective and concurrent A-to-I and C-ot-U RNA base editing

Anna S. Stroppel, **Ngadhnjim Latifi**, Alfred Hanswillemenke, Rafail Nikolaos Tasakis, F. Nina Papavasiliou and Thorsten Stafforst, *Nucleic Acids Research*, 2021, *49.16,* e95

Harnessing self-labeling enzymes for selective and concurrent A-to-I and C-to-U RNA base editing

Anna S. Stroppel[1], Ngadhnjim Latifi[1], Alfred Hanswillemenke[1], Rafail Nikolaos Tasakis[2,3], F. Nina Papavasiliou[2,3] and Thorsten Stafforst[1,*]

[1]Interfaculty Institute of Biochemistry, University of Tübingen, Auf der Morgenstelle 15, 72076 Tübingen, Germany, [2]Division of Immune Diversity (D150), German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany and [3]Faculty of Biosciences, University of Heidelberg, 69120 Heidelberg, Germany

Received November 03, 2020; Revised May 05, 2021; Editorial Decision June 07, 2021; Accepted June 18, 2021

ABSTRACT

The SNAP-ADAR tool enables precise and efficient A-to-I RNA editing in a guideRNA-dependent manner by applying the self-labeling SNAP-tag enzyme to generate RNA-guided editases in cell culture. Here, we extend this platform by combining the SNAP-tagged tool with further effectors steered by the orthogonal HALO-tag. Due to their small size (ca. 2 kb), both effectors are readily integrated into one genomic locus. We demonstrate selective and concurrent recruitment of ADAR1 and ADAR2 deaminase activity for optimal editing with extended substrate scope and moderate global off-target effects. Furthermore, we combine the recruitment of ADAR1 and APOBEC1 deaminase activity to achieve selective and concurrent A-to-I and C-to-U RNA base editing of endogenous transcripts inside living cells, again with moderate global off-target effects. The platform should be readily transferable to further epitranscriptomic writers and erasers to manipulate epitranscriptomic marks in a programmable way with high molecular precision.

INTRODUCTION

After transcription, most RNA species get processed (e.g. capped, spliced, trimmed, polyadenylated) and enzymatically modified (1). Particularly wide-spread modifications include methylation (e.g. m^6A, 2′-*O*-methylation), isomerization (pseudouridine) and deamination (e.g. A-to-I and C-to-U editing). Due to recent progress in deep sequencing technologies, the fundamental role of such epitranscriptomic modifications in human pathophysiology became apparent (2,3), including the biology of learning (4), development (5) and cancer (6,7). A detailed mechanistic understanding of the plethora of epitranscriptomic modifications is currently hampered by a lack of methods to manipulate transcripts in a programmable way with molecular precision (8). Fortunately, RNA transcripts are precisely addressable via Watson-Crick base pairing. Thus, a guideRNA can be applied to recruit a protein effector to a specific transcript in a site-specific manner. During the last years, various attempts focused on the engineering of RNA-guided RNA base editing effectors, specifically on A-to-I and C-to-U editing (8). As inosine is biochemically interpreted as guanosine, site-directed RNA editing enables the reprogramming of genetic information, e.g. substitution of amino acids, formation and removal of premature termination codons, which open novel avenues for drug discovery, promising to bypass technical and ethical issues related to genome editing (8). In this regard, our group developed an RNA-targeting platform based on fusion proteins of the self-labeling SNAP-tag (Figure 1A). To engineer a programmable A-to-I RNA base editor, we fused the SNAP-tag with the catalytic domain of the RNA editing enzyme ADAR (9,10), more specifically, we have used a hyperactive mutant (11), carrying a single glutamate (E) to glutamine (Q) mutation, indicated by the letter Q. In these fusions, the SNAP-tag (12) exploits its self-labeling activity to covalently attach to a guideRNA in a defined 1:1 stoichiometry by recognizing a benzylguanine (BG) moiety at the guideRNA (13). The guideRNA then addresses the editing of one specific adenosine residue in a selected transcript with high efficiency, broad codon scope, and very good precision (9). Competing RNA-targeting platforms, e.g. based on Cas proteins (14,15) or tethering approaches, have been developed for similar applications (8,10,16,17). Each approach has different strengths and weaknesses (8,10). A clear advantage of the SNAP-tag approach is its human origin, the small size, the ease of stable expression, the ease of transfecting one or multiple chemically stabilized guideRNA(s), which allows for concurrent editing (9), and the ready inclusion of photo control (18,19). Here, we extend the self-labeling RNA-targeting platform with HALO-tag fusions and characterize their abilities to recruit two different editing effectors in an orthogonal fash-

*To whom correspondence should be addressed. Tel: +49 7071 29 75376; Email: thorsten.stafforst@uni-tuebingen.de

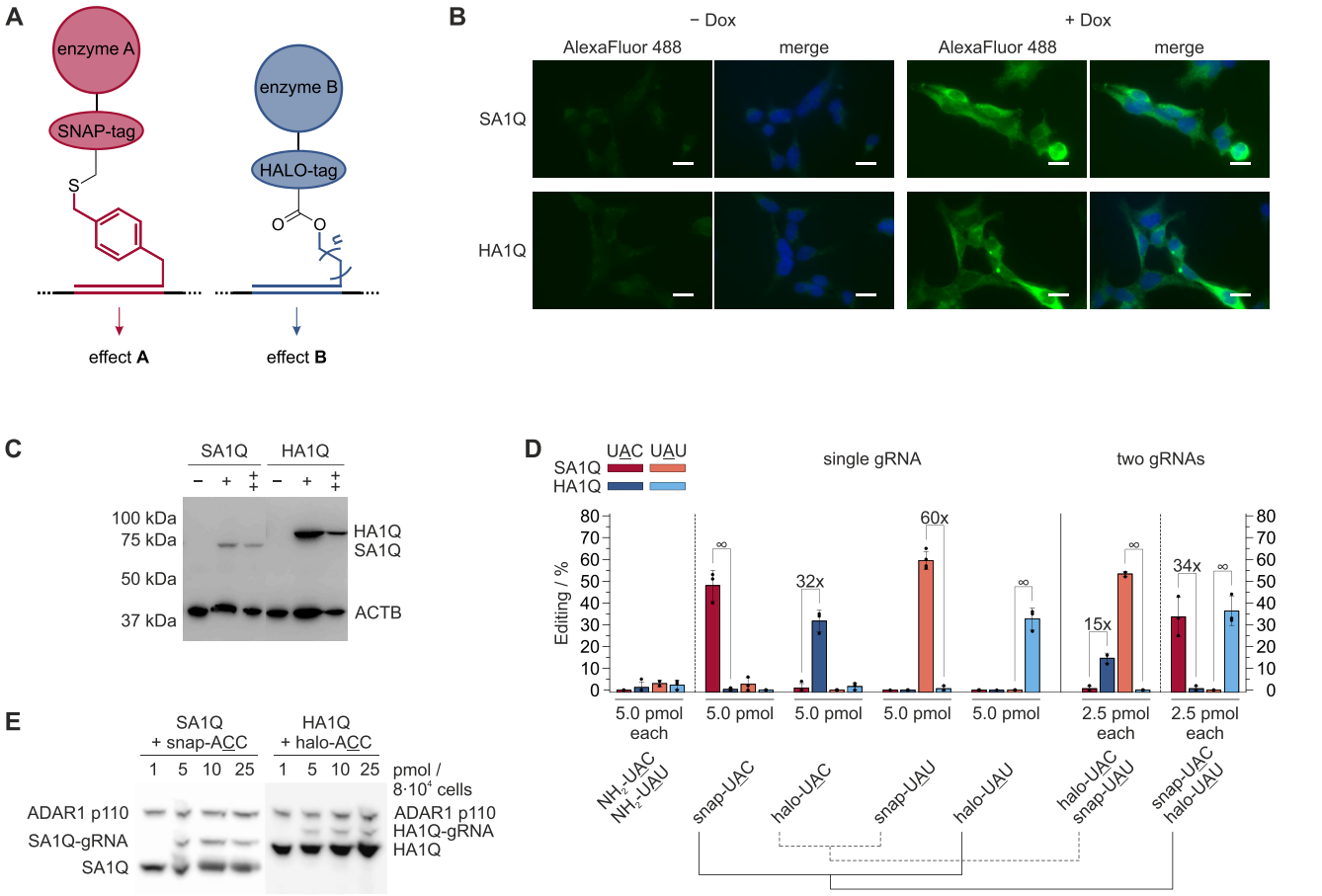


Figure 1. Recruitment of the ADAR1 deaminase domain in fusion with two different self-labeling enzymes. (**A**) Independent self-labeling enzymes, e.g. SNAP- and HALO-tag, enable for the orthogonal recruitment of various effectors, e.g. enzymes A and B. (**B**) Characterization of 293 Flp-In T-REx cell lines expressing either the Myc-tagged SA1Q or HA1Q transgene in a doxycycline-dependent fashion as visualized by immunostaining with α-Myc (green channel) and DNA staining with Hoechst 33342 (blue channel). Scale bars correspond to 15 μm. (**C**) Western blot (α-Myc) to compare SA1Q and HA1Q expression. + means 24 h, ++ means 48 h doxycycline induction. (**D**) Editing efficiency and orthogonality of four different guideRNAs (snap-UAC, halo-UAC, snap-UAU, halo-UAU) targeting either a 5′-UAC or 5′-UAU codon in the ORF of endogenous GAPDH. Either single guideRNAs (left panel) or the indicated combination of two guideRNAs (right panel) were transfected into the SA1Q or HA1Q cell line, as indicated in the legend respectively. NH_2-guideRNAs are control guideRNAs with same sequence but lacking a self-labeling moiety. Data are shown as the mean ± SD of $N = 3$ independent experiments. (**E**) Dose-dependent formation of SA1Q- and HA1Q-guideRNA conjugates (SA1Q-gRNA and HA1Q-gRNA) after transfection of 1.0, 5.0, 10 or 25 pmol snap- or halo-guideRNA per 8×10^4 cells respectively, visualized via Western blot (α-ADAR1). Endogenous ADAR1 p110 is equally expressed independent of guideRNA addition.

ion (Figure 1A). This broadens the otherwise limited codon scope of single editing enzymes, and enables site-selective, concurrent A-to-I and C-to-U editing within the same cell.

MATERIALS AND METHODS

Reagents and biological resources

Detailed information on reagents, enzymes, antibodies and kits as well as cell lines used in this study are presented in the Supporting Information.

Chemical synthesis

The self-labeling moieties that were attached to the guideRNAs, i.e. snap, clip, halo, halo-snap, $(snap)_2$ and $(halo)_2$ were synthesized via solid phase peptide synthesis as described in the Supporting Information (Supplementary Schemes S1–S3, Supplementary Figures S1–S5).

Generation of guideRNAs

As guideRNAs, 22 nt long RNAs with a 5′-C6-aminolinker (NH_2-guideRNAs) that were chemically stabilized in an antagomir-like fashion as described before (20) were applied. Additional details as well as sequences and extinction coefficients at 260 nm of all used guideRNAs can be found in the Supporting Information (Supplementary Table S1).

snap-, clip- and halo-guideRNAs were produced analogous to the previously reported protocol for Npom-guideRNAs (18). Instead of N^7-Npom-BG-Linker-COOH, 8 μl (60 mM in DMSO, 480 nmol, ∼35 eq) of either snap, clip or halo were used. snap- and clip-guideRNAs were purified via precipitation as described before (18). For halo-guideRNAs, samples were lyophilized after aqueous extrac-

tion from the urea PAGE and subsequently purified with C18 Reversed Phase Cartridges (WATERS, #020515) according to manufacturer's manual.

halo-snap-, $(snap)_2$- and $(halo)_2$-guideRNAs were produced analogous to the previously reported improved protocol with DIC activation (21), using 4 μl (60 mM in DMSO, 240 nmol, ∼17.5 eq) of either halo-snap, $(snap)_2$ or $(halo)_2$. $(snap)_2$-guideRNAs were purified via precipitation as described before (21), halo-snap- and $(halo)_2$-guideRNAs were again purified with C18 Reversed Phase Cartridges (WATERS, #020515) according to manufacturer's manual.

Generation of stable cell lines

In general, cells were cultivated in Dulbecco's modified Eagle's medium (DMEM, LIFE TECHNOLOGIES) supplemented with 10% fetal bovine serum (FBS, LIFE TECHNOLOGIES) at 37 °C with 5% CO_2 in a water saturated steam atmosphere. For generating stable, inducible cell lines, the Flp-In™ T-REx™ system by LIFE TECHNOLOGIES was used. 4×10^6 293 Flp-In T-REx cells were seeded in 10 ml DMEM/10% FBS/100 μg/ml zeocin/15 μg/ml blasticidin (DMEM/FBS/Z/B) in a 10 cm dish. After 23 h, medium was replaced with DMEM/10 % FBS (DMEM/FBS) and 1 h later 9 μg pOG 44 and 1 μg of the respective construct in a pcDNA 5 vector were forward transfected with 30 μl Lipofectamine 2000 (THERMO FISHER SCIENTIFIC). After 24 h, medium was replaced with 15 ml DMEM/10% FBS/15 μg/ml blasticidin/100 μg/ml hygromycin (DMEM/FBS/B/H), followed by selection for approximately two weeks. Then, the stable cell lines were transferred to a 75 cm^2 cell culture flask and subsequently cultivated in DMEM/FBS/B/H. Sequences of the constructs for all cell lines used in this study can be found in the Supporting Information.

Immunostaining of single cell lines

Briefly, 1.2×10^5 SA1Q or HA1Q 293 Flp-In T-REx cells were seeded on coverslips coated with poly-D-lysine in DMEM/FBS/B/H for –Dox samples or DMEM/FBS/B/H/10 ng/ml doxycycline (DMEM/FBS/B/H/10 D) for +Dox samples respectively. After 24 h, cells were fixed with 3.7% formaldehyde in PBS, permeabilized with 1% Triton X-100 in PBS and blocked with 10% FBS in PBS. Cells were then incubated with mouse α-Myc (1:1000 in 10% FBS in PBS, SIGMA ALDRICH M4439), followed by goat α-mouse Alexa Fluor 488 (1:1000 in 10% FBS in PBS, THERMO FISHER SCIENTIFIC A11001). Nuclei were stained with NucBlue™ Live ReadyProbes™ Reagent Hoechst33342 (1:100 in PBS, THERMO FISHER SCIENTIFIC R37605) and coverslips were mounted to object slides with Fluorescence Mounting Medium by DAKO. Microscopy was performed with a ZEISS AXIO Observer.Z1 with a Colibri.2 light source under 63× magnification. For further procedural details, excitation and emission wavelengths, see Supporting Information (Supplementary Table S2).

FITC-BG & TMR-chloroalkane staining of duo cell lines

5×10^4 293 Flp-In T-REx cells from cell lines **1**–**5** were seeded on coverslips coated with poly-D-lysine in DMEM/FBS/B/H for –Dox samples or DMEM/FBS/B/H/10 D for +Dox samples respectively. After 24 h, cells were stained with 2 μM FITC-BG, 5 μM TMR-chloroalkane and NucBlue™ Live ReadyProbes™ Reagent Hoechst33342 (1:100, THERMO FISHER SCIENTIFIC R37605). Cells were then fixed with 3.7% formaldehyde in PBS, permeabilized with 0.1% Triton X-100 in PBS and coverslips were mounted to object slides with Fluorescence Mounting Medium by DAKO. Microscopy was performed with a ZEISS AXIO Observer.Z1 with a Colibri.2 light source under 63× magnification. For experimental data of –Dox samples, further procedural details, excitation and emission wavelengths, see Supporting Information (Supplementary Figure S6, Table S2).

Western blotting of protein expression in single cell lines

Briefly, 1×10^5 SA1Q or HA1Q 293 Flp-In T-REx cells respectively were seeded and treated with doxycycline for 24 h (+) or 48 h (++) or left uninduced (–). After 48 h, cells were harvested and lysed in urea lysis buffer (8 M urea, 100 mM NaH_2PO_4, 10 mM Tris, pH 8.0) via shear force. Protein lysates were separated via SDS-PAGE and transferred onto a PVDF membrane (BIO-RAD LABORATORIES). After blocking in 5% dry milk in TBST containing 50 μg/ml avidin, the blot was incubated with mouse α-Myc (1:5000, SIGMA ALDRICH M4439) and mouse α-ACTB (1:40 000, SIGMA-Aldrich A5441) in 5% dry milk-TBST as primary antibodies. As secondary antibody, goat α-mouse HRP (1:10 000, JACKSON IMMUNORESEARCH 115-035-003) with added Precision Protein StrepTactin HRP conjugate (for visualisation of the Precision Plus Western C Standard, 1:25 000, BIO-RAD) in 5% dry milk-TBST was applied. Chemiluminescence was measured with a FUSION FX by VILBER. For full Western Blot and further experimental details, see Supporting Information (Supplementary Figure S7).

Western blotting of guideRNA–protein conjugation

2×10^6 SA1Q or HA1Q 293 Flp-In T-REx cells were seeded in DMEM/FBS/B/H/10 D. After 24 h, 4×10^5 cells were reverse transfected with the respective amount of snap- or halo-ACC with 2.5 μl Lipofectamine 2000. Doxycycline concentration was kept at 10 ng/ml and after further 24 h cells were lysed in 1× Laemmli (67 mM SDS, 10 mM Tris pH 6.8, 1.1 M glycerol, 0.10 M dithiothreitol, 0.15 mM bromophenol blue) in RIPA Lysis and Extraction Buffer (1% NP-40, 150 mM NaCl, 25 mM Tris·HCl pH 7.6, 1% sodium deoxycholate, 0.1% SDS, THERMO FISHER SCIENTIFIC; supplemented with 1 tablet cOmplete™ Mini EDTA-free Protease Inhibitor Cocktail by ROCHE per 10 ml). Protein lysates were separated via SDS-PAGE and transferred onto a PVDF membrane (BIO-RAD LABORATORIES). After blocking in 5% dry milk in TBST, the blot was incubated with rabbit α-ADAR1 (1:1000, BETHYL LABORATORIES A303-884) and rabbit α-GAPDH (1:1000, CELL SIGNALING #5174) in 5% dry milk-TBST as primary antibodies. As

secondary antibody, goat α-rabbit HRP (1:10 000, JACKSON IMMUNORESEARCH 111-035-003) in 5% dry milk-TBST was applied. Chemiluminescence was measured with an Odyssey Fc Imaging System (LI-COR). For additional experimental data as well as further procedural details, see Supporting Information (Supplementary Figure S8).

TMR-staining & western blotting of protein expression in duo cell lines

2×10^5 293 Flp-In T-REx cells from the respective duo cell line were seeded in DMEM/FBS/B/H for –Dox samples or DMEM/FBS/B/H/10 D for +Dox samples respectively. After 24 h, cells were harvested and lysed in NP40 lysis buffer (1% NP-40, 150 mM NaCl, 50 mM Tris pH 8.0; 1 tablet cOmplete™ Mini EDTA-free Protease Inhibitor Cocktail by ROCHE per 10 ml). For co-staining with TMR-BG and TMR-chloroalkane, protein lysate was incubated with 5 μM TMR-BG and TMR-chloroalkane each in NP40 lysis buffer for 30 min at 37°C and 600 rpm. Protein lysates were then separated via SDS-PAGE and TMR-staining was visualized on a FLA 5100 by FUJIFILM with excitation at 532 nm and emmission at 557 nm (Cy3 filter set). Subsequently, proteins were transferred onto a PVDF membrane (BIO-RAD LABORATORIES), and the blot was blocked in 5% dry milk in TBST containing 50 μg/ml avidin, followed by incubation with mouse α-ACTB (1:40 000, SIGMA-Aldrich A5441), rabbit α-SNAP-tag (1:1000, NEW ENGLAND BIOLABS P9310S) and rabbit α-HaloTag (1:1000, PROMEGA G9281) in 5% dry milk-TBST as primary antibodies. As secondary antibodies, goat α-mouse HRP (1:5000, JACKSON IMMUNORESEARCH 115-035-003) with added Precision Protein StrepTactin HRP conjugate (for visualisation of the Precision Plus Western C Standard, 1:25 000, BIO-RAD) and goat α-rabbit HRP (1:5000, JACKSON IMMUNORESEARCH 111-035-003) were applied. Chemiluminescence was measured with a FUSION FX by VILBER. For additional experimental data as well as further procedural details, see Supporting Information (Supplementary Figure S9).

Editing of endogenous targets

For the editing experiments, 4×10^5 of the respective 293 Flp-In T-REx cells were seeded in DMEM/FBS/B/H/10 D. After 24 h, 8×10^4 cells were reverse transfected with the respective amount of the guideRNA to be examined with 0.5 μl Lipofectamine 2000. Doxycycline concentration was kept at 10 ng/ml and after further 24 h (or 48 h for cell lines expressing APO1S) cells were harvested. RNA isolation was performed with the Monarch® RNA cleanup kit from NEW ENGLAND BIOLABS, followed by DNase I digestion. Samples containing $(snap)_2$-A$\underline{C}$C were treated with a DNA oligonucleotide of complementary sequence (anti-$(snap)_2$-A$\underline{C}$C, 1 μM) at 95°C for 3 min to trap the guideRNA. Purified RNA was then reverse transcribed to cDNA, which was amplified via Taq PCR and subsequently analyzed with Sanger sequencing (either EUROFINS GENOMICS or MICROSYNTH). A-to-I editing yields were determined by dividing the peak height for guanosine by the sum of the peak heights for both adenosine and guanosine. Additional experimental data and further procedural details are given in the Supporting Information (Supplementary Figures S12 and S13, Supplementary Table S3).

Editing of transfected reporter transcript

For editing of the reporter transcript, cells were forward transfected 24 h after seeding with 300 ng pcDNA 3.1 containing the coding sequence for eGFP-W58X with 1.2 μl Lipofectamine 2000. 24 h thereafter, 8×10^4 cells were reverse transfected with the respective amount of the guideRNA to be examined with 0.5 μl Lipofectamine 2000. Cells were harvested after further 48 h and proceeded as for editing of endogenous targets. For additional experimental data and procedural details, see Supporting Information (Supplementary Figure S14).

Next generation sequencing

For cell line **2** and **9**, four samples each were prepared for NGS, i.e. a duplicate of an empty transfection and a duplicate of a guideRNA transfection (0.5 pmol $(snap)_2$-C$\underline{A}$G and $(halo)_2$-C$\underline{A}$U for cell line **2**, 2.5 pmol $(halo)_2$-U$\underline{A}$U and $(snap)_2$-A$\underline{C}$C for cell line **9**), all under doxycycline induction. RNA was isolated, DNase I digested and purified via RNeasy MinElute Cleanup Kit from QIAGEN. mRNA next generation sequencing was then performed by CEGAT. The library was prepared with the library preparation kit TruSeq Stranded mRNA by ILLUMINA starting from 100 ng RNA. Samples were then sequenced on a NovaSeq 6000 by ILLUMINA with 50 million reads and 2 × 100 bp paired end. RNA-seq raw data from different lanes that belong to the same sample were pulled together. After adapter trimming with Trim Galore (v. 0.6.4; http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), the trimmed reads were aligned using STAR (v. 2.7.3a) (22) to a genome index inferred by the human reference genome (hg19) sequence, along with the RefSeq annotation, both publicly available at the genome browser at UCSC (23). For the alignments we considered reads that were uniquely mapped (STAR option: –outFilterMultimapNmax 1) to avoid multimapping between highly similar regions. Aligned data (bam files) were deduplicated, sorted and indexed with SAMtools (v. 1.9; http://samtools.sourceforge.net) (24). SNVs in our samples were called with REDItools (v2; https://github.com/tflati/reditools2.0) (25,26), considering the developers' recommendations for data preparation prior to this step. Sticking to our previously published approach (9), we considered only high-quality sites (min. MeanQ > 30 in REDItools2), and we called editing in well-covered sites (min. 50 reads in aggregate of the two replicates per sample) that showed ≥10% (for A-to-I) or ≥ 5% (for C-to-U) editing frequency when compared to the control. Additionally, fisher's exact tests were performed for all the sites that fulfilled the aforementioned criteria and significantly differentially edited sites were considered those that showed adjusted P-value <0.01. Sites that were reported in the first 6 sites of a read, or in homopolymeric regions, or reported in the dbSNP (v. 142; excluding cDNA-based reported SNPs: http://www.ncbi.nlm.nih.gov/SNP/), were ex-

cluded throughout our output lists. All genomic coordinates were annotated with Oncotator (v1.9.9.0) (27) and Repeat Mask for Alu-SINE elements of UCSC Genome Browser (23) both for hg19. Additional data, including scatter plots of total off-targets in all editing experiments, elaborate analysis of significantly differently edited sites with editing difference ≥25%, analysis of bystander editing sites and scatter plots of all called editing sites in the two respective replicates, as well as details on the experimental procedure can be found in the Supporting Information (Supplementary Figures S16–S23, Supplementary Tables S6–S12).

RESULTS AND DISCUSSION

The HALO-tag outperforms the CLIP-tag to complement the RNA targeting platform

Two self-labeling enzymes are to be considered to complement the SNAP-tag for RNA targeting, the HALO-tag (28) and the CLIP-tag (29). The HALO-tag covalently attaches to halo-guideRNAs, carrying a 1-chloroalkane moiety (28), the CLIP-tag to clip-guideRNAs, carrying a benzylcytosine moiety for covalent conjugation (29), both in 1:1 stoichiometry. In a preliminary experiment, we identified the HALO-tag as the preferred tag for two reasons. First, a clip-guideRNA gave notable editing also with SNAP-ADAR, indicating insufficient orthogonality (29) between SNAP- and CLIP-tag in the editing application (Supplementary Figures S10–S12). Second, the clip-guideRNA showed loss of activity upon long-term storage (Supplementary Figure S12). We thus continued to compare HALO-ADAR1 (HA1Q) with SNAP-ADAR1 (SA1Q), our best RNA editor from our previous study (9). Both fusions carried the hyperactive Q mutation in the deaminase domain. Plasmid overexpression of editing enzymes typically results in enormous variability of expression levels, massive off-target editing, and low and unsteady editing efficiency at endogenous targets (10). To avoid such artefacts, we generated cell lines stably expressing either HA1Q or SA1Q from a defined, single genomic site, under control of doxycycline, by applying the 293 Flp-In T-REx system (9,19). Both cell lines expressed the respective fusion protein in a homogenous and doxycycline-inducible manner (Figure 1B). Both fusions were localized in nucleoplasm and cytoplasm, favoring the latter. The expression level of HA1Q was slightly higher compared to SA1Q (Figure 1C).

Snap- and halo-guideRNAs recruit SNAP- and HALO-fusions with high selectivity

To examine editing efficiency and orthogonality, we generated four guideRNAs and transfected them separately either into the HA1Q or SA1Q cell line. Two guideRNAs were designed to target a 5′-UAC codon in the ORF of GAPDH and were only differing in the self-labeling moiety, being either benzylguanine (12) (snap-UAC) for SNAP-tag or chloroalkane (28) (halo-UAC) for HALO-tag conjugation. Another pair of guideRNAs was equally designed to target a 5′-UAU codon in GAPDH. We observed very selective and orthogonal editing, both snap-guideRNAs elicited editing only in the SA1Q cell line as both halo-guideRNAs did in the HA1Q cell line (Figure 1D, left panel). Furthermore, editing was reliably programmable and editing in the non-targeted codon was not observed. Even though slightly higher expressed, HA1Q was less active than SA1Q on both targets. We checked the in situ assembly of each fusion protein with its respective guideRNA by Western blot (Figure 1E). Both couples gave a similar dose-dependent formation of the protein–guideRNA conjugate not exhausting the protein component at guideRNA amounts typically applied in editing reactions. Thus, neither expression level nor conjugation efficiency explains the slightly reduced editing efficiency of HA1Q. Co-transfection of two guideRNAs, one halo- and one snap-guideRNA, gave decent editing with high selectivity for the matching enzyme in each respective cell line (Figure 1D, right panel), highlighting that the co-transfection of a guideRNA with mismatching self-labeling moiety is possible and does not interfere with the selectivity of the matching guideRNA.

Cell lines co-expressing SNAP- and HALO-tagged effectors are easily generated

Next, we explored the selective and concurrent recruitment of two different effectors based on the orthogonal self-labeling reactions mediated by SNAP- and HALO-tag within one cell (Figure 1A). As effectors, we first combined two different A-to-I RNA editing enzymes, and later one A-to-I with one C-to-U RNA editase.

ADAR1 and ADAR2 have partly complementing substrate preferences (9,30). Hence, their orthogonal recruitment inside a cell is highly desired and we decided to co-express the newly characterized HA1Q (Figure 1) with the formerly characterized (9) SA2Q. In contrast to competing RNA targeting platforms, e.g. based on Cas proteins, self-labeling proteins are of small size with only 2.2 kb for HA1Q and 1.8 kb for SA2Q. This enabled us to generate small co-expression cassettes in the pcDNA 5 backbone which allow for their targeted integration into the FRT recombination site of 293 Flp-In T-REx cells (9,19). The strong expression of two transgenes within close proximity often leads to their mutual transcriptional interference (31). Thus, we constructed five different cassettes (Figure 2A), varying the relative positioning of the two transgenes, their promotors (CMV or Ef1α), and their direction of transcription. We also tested a P2A (32) fusion construct that drives both transgenes from one promotor. All five constructs were integrated into the 293 Flp-In T-REx parent cell line by simple plasmid transfection to generate duo cell lines that express both transgenes homogenously among the cell population under doxycycline control (Figure 2B, Supplementary Figure S6). Importantly, ready-to-use duo cell lines were obtained after two weeks of antibiotic selection with no need for cumbersome clonal selection. To better characterize the relative transgene expression in duo cell lines **1**–**5**, we stained both HA1Q and SA2Q in a defined 1:1 stoichiometry with tetramethylrhodamine (TMR) by adding TMR-benzylguanine and TMR-chloroalkane to full cell lysate and analyzed the stained proteins after SDS-PAGE separation (Figure 2C, Supplementary Figure S9). In a preliminary editing experiment, we tested for the editing activity of both transgenes in all five duo cell lines and found

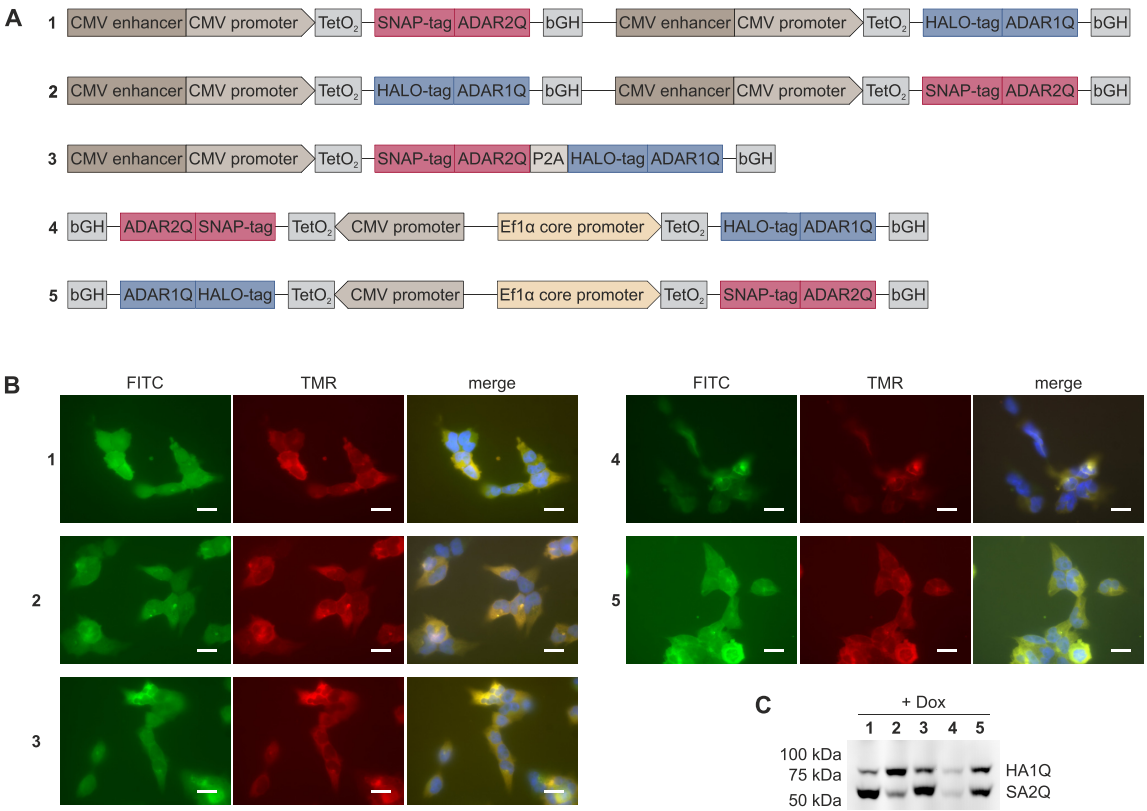


Figure 2. Generation of duo cell lines 1–5 for homogenous co-expression of two transgenes. (**A**) Constructs (**1**–**5**) were designed to co-express both transgenes (HA1Q and SA2Q) from one cassette under doxycycline control. $TetO_2$: tet operator, leads to repression of expression in the absence of a tetracycline (33); bGH: bovine growth hormone terminator; P2A: porcine teschovirus-1 self-cleaving 2A peptide (32). (**B**) All duo cell lines have been characterized for the transgene co-expression by staining with FITC-BG (green channel) and TMR-chloroalkane (red channel). Cell nuclei are stained with Hoechst 33342 (blue channel). Scale bars correspond to 15 μm. (**C**) Characterization of relative transgene expression via SDS-PAGE after co-staining with TMR-BG and TMR-chloroalkane in raw cell lysate.

HA1Q expression to be the major limiting factor (Supplementary Table S3). We continued the study largely based on duo cell line **2**, which expressed HA1Q to the highest level and SA2Q to a level sufficient to obtain good editing yields.

Selective recruitment of ADAR1 and ADAR2 activity extends the codon scope

ADAR1 and ADAR2 partly prefer different codons (34,35). We have comprehensively characterized the codon preferences of SA1Q and SA2Q before (9) and found, for example, that the 5′-CAG codon was preferentially edited by SA2Q, with a 3.3-fold higher editing yield compared to SA1Q, while the 5′-CAU codon was preferentially edited by SA1Q, with a 6.3-fold higher editing yield (Figure 3A). Thus, a cell line expressing only one of the two RNA base editors will not permit optimal editing yields in any case. In contrast, we predict that the selective recruitment of HA1Q and SA2Q with halo- and snap-guideRNAs, will enable to recruit the preferred enzyme to any substrate (matching combination, Figure 3B). Accordingly, we can predict the existence of a mismatching combination of guideRNAs that will lead to inferior editing results on both targets.

Initially, we tested this by transfection of single guideRNAs into duo cell line **2** (Figure 3C, left panel). GuideRNAs were either targeting a 5′-CAG codon in the ORF of ACTB or a 5′-CAU codon in the ORF of GAPDH. Furthermore, guideRNAs were either equipped with a BG moiety (snap-guideRNA) or with a chloroalkane moiety (halo-guideRNA) to selectively recruit SA2Q or HA1Q, respectively. Indeed, recruitment of SA2Q with the snap-CAG guideRNA always gave better editing yields for the 5′-CAG codon in ACTB than recruitment of HA1Q with the halo-CAG guideRNA. As expected, the effect was reverse for the editing of the 5′-CAU codon in GAPDH. Notably, only the halo-CAU guideRNA, selective for HA1Q, was able to induce detectable editing at all. A strength of the SNAP-ADAR platform is the ease by which the short (ca. 20 nt), chemically modified guideRNAs can be transfected into cells. In the past, we demonstrated co-transfection of up to four different guideRNAs enabling multiplexed, concurrent editing of four different substrates without loss in editing efficiency (9). Now, we co-transfected two guideRNAs, one snap- and one halo-guideRNA, either in matching or mismatching combination into cell line **2**. Clearly, the matching combination gave better editing yields for both substrates (CAG, CAU) compared to the mismatching combination. Again,

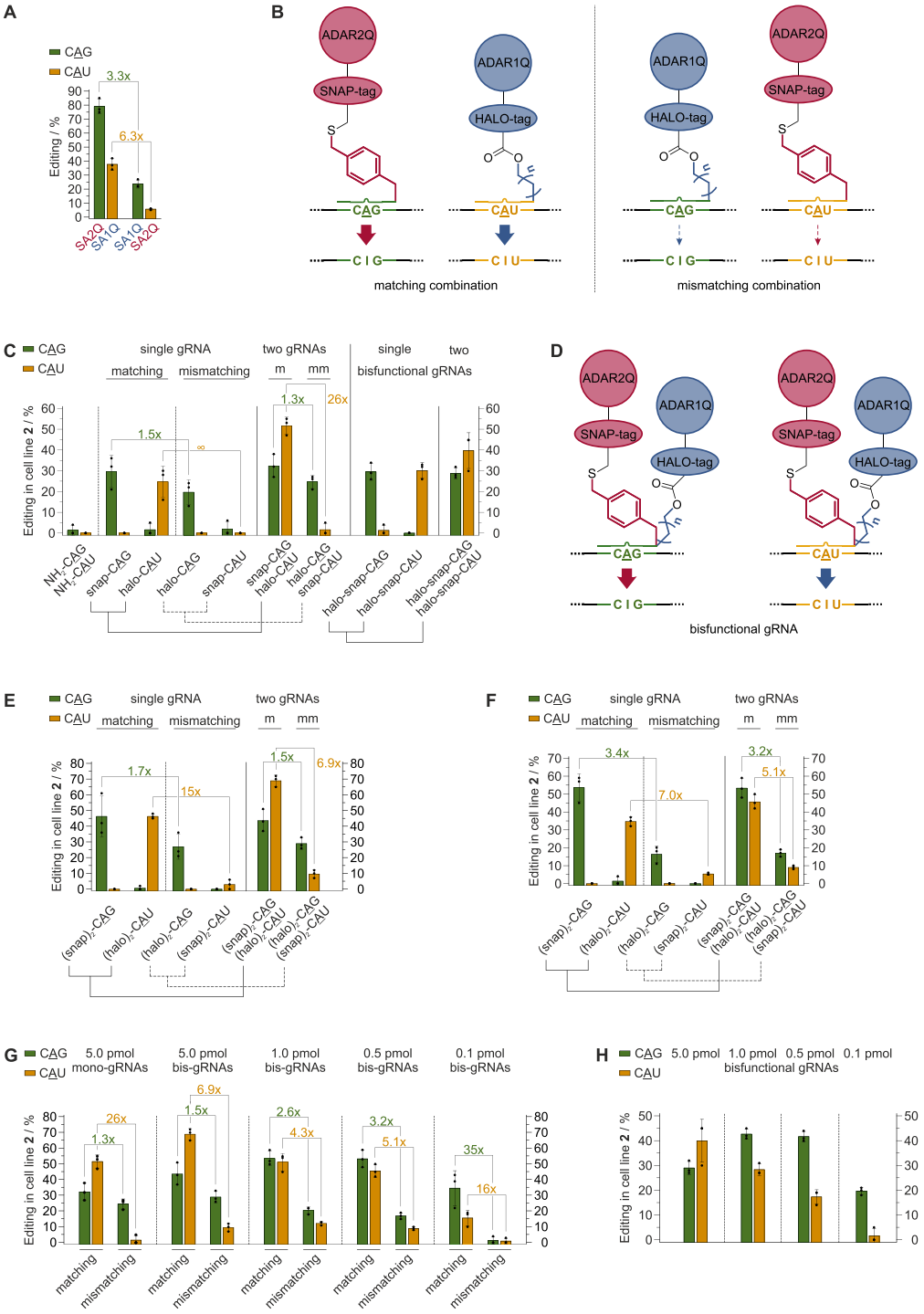


Figure 3. Editing in duo cell lines expressing HA1Q and SA2Q. (**A**) SA1Q and SA2Q have different preferences for 5′-CAG and 5′-CAU codons in the ORF of GAPDH, as described before (9). (**B**) Due to the two different self-labeling moieties (BG and chloroalkane) the SNAP-tagged ADAR2 and the HALO-tagged ADAR1 deaminase domains can be recruited either to their preferred substrates (matching combination) or to their least preferred substrates (mismatching combination). (**C**) Editing yield and selectivity after transfection of a single (5.0 pmol), matching or mismatching snap- or halo-guideRNA into duo cell line **2** compared to the co-transfection of two guideRNAs (one snap- and one halo-guideRNA, each 5.0 pmol) either in matching (m) or in mismatching (mm) combination (left panel). The right panel shows the activity of bisfunctional guideRNAs capable to recruit both editing enzymes with one guideRNA. (**D**) Bisfunctional halo-snap-guideRNAs, carrying both a chloroalkane and a BG moiety, are able to recruit both HA1Q and SA2Q, leading to maximum editing yields at any codon. (**E**) Editing yield and selectivity in duo cell line **2** after transfection of a single or co-transfection of two guideRNAs, one $(snap)_2$- and one $(halo)_2$-guideRNA, either in matching (m) or in mismatching (mm) combination (5.0 pmol each). (**F**) Same as E) but with 0.5 pmol each. (**G**) Concentration dependency of editing efficiency and selectivity in cell line **2** under co-transfection of $(snap)_2$- and $(halo)_2$-guideRNAs (bis-guideRNAs) in matching versus mismatching combination. For comparison, editing with the respective mono-guideRNAs (snap- and halo-guideRNAs) is shown. (**H**) Concentration dependency of editing yields in duo cell line **2** after co-transfection of two bisfunctional halo-snap-guideRNAs. Data in a), c), e)-h) are shown as the mean ± SD of $N = 3$ independent experiments.

choosing the matching combination was required to see editing with the CAU substrate at all. The same pattern was observed for a second duo cell line, cell line **5** (Supplementary Figure S13a). This demonstrates that the platform is able to target two editing enzymes independently from each other to their respective preferred target inside one cell line.

One could also conceive a bisfunctional guideRNA capable of recruiting both editases, HA1Q and SA2Q, simultaneously (Figure 3D). Such a halo-snap-guideRNA may enable maximum editing with any codon and substrate. To accomplish that, we synthesized halo-snap-guideRNAs carrying both, the BG and the chloroalkane moiety, targeting either the CAG or CAU substrate and tested them in duo cell line **2**. As expected, both halo-snap-guideRNAs gave good editing yields for both codons, 5′-CAU and 5′-CAG, always resembling the editing result of the formerly preferred snap- or halo-guideRNA, respectively (Figure 3C, right panel). This clearly indicates that both enzymes have been active on the substrates.

As controls, we had also synthesized $(snap)_2$- and $(halo)_2$-guideRNAs carrying either two benzylguanine or two chloroalkane moieties, respectively. Notably, editing yields have been higher with such controls (Figure 3E, F) compared to the respective guideRNAs carrying only one self-labeling moiety. This boost might be due to the recruitment of two instead of one editing enzyme per guideRNA. Similar effects have been described in the context of other RNA editing systems before (36). Interestingly, not only the yield but also the selectivity (e.g. CAG codon) was better than before (Figure 3F). One can expect that the selectivity increases further if one reduces the concentration of the guideRNA-enzyme conjugate inside the cell. Thus, we varied the amount of the two transfected guideRNAs (one $(snap)_2$- and one $(halo)_2$-guideRNA, either matching or mismatching) between 5 pmol and 0.1 pmol in four steps (Figure 3G, Supplementary Figure S13b, c). Indeed, stepwise reduction of the guideRNA amount improved the selectivity progressively. At 0.1 pmol guideRNA, excellent selectivity was obtained with virtually no residual editing on both targets (CAG and CAU) in the mismatching combination. Notably, the editing yields were satisfying also at low amounts of guideRNA. A similar trend, but with lower editing yields, was seen for the bisfunctional halo-snap-guideRNAs (Figure 3H, Supplementary Figure S13d) indicating that the recruitment of two copies of the preferred editing enzyme gives better editing yield than the co-recruitment of one preferred and one non-preferred enzyme.

Genomic co-expression of two editing enzymes elicits moderate global off-target editing

Overexpression of engineered, highly active editing enzymes leads to significant off-target editing throughout the whole transcriptome (8,10). Various strategies have been tried to minimize this (8,10). In this regard, we demonstrated that the controlled expression of SA1Q and SA2Q from single genomic loci reduces global off-target editing tremendously (9). We now determined the total off-target editing in duo cell line **2** after co-transfection with 0.5 pmol $(snap)_2$-CAG and 0.5 pmol $(halo)_2$-CAU guideRNA, by determining significantly differently edited sites in comparison with a negative control expressing no artificial editing enzyme. As the pipeline was more sensitive than the one used before (9), we re-analyzed the raw data of the total off-target editing for mono cell lines expressing SA1Q or SA2Q, in presence of an ACTB-targeting snap-guideRNA (9), with the new pipeline to allow for direct side-by-side comparison with duo cell line **2**. With 8391 sites, the amount of total off-target editing in duo cell line **2** roughly comprised the aggregate of sites found in mono cell lines SA1Q and SA2Q (Table 1, Figure 4A, Supplementary Figure S16). However, the vast majority of editing sites (ca. 75%) showed changes in editing levels below 25% (Supplementary Table S6, Figure 4A). The total off-targets comprise guideRNA-dependent and -independent editing events. To determine the guideRNA-dependent fraction we compared the off-target editing for cell line **2** with versus without co-transfection of the two guideRNAs. Our sensitive pipeline detected 653 sites that were significantly differently edited depending on the presence of the guideRNAs (Figure 4B, Table 1). Again, only a small number of sites (Supplementary Table S6) showed editing sites with levels elevated above 25%. Among these 37 sites, only five sites were missense mutations. After careful analysis, almost all 37 sites could be assigned to either binding of the GAPDH or ACTB guideRNA, respectively (Supplementary Figures S17–S19). Notably, only one missense mutation (ACTA2, 47%) achieved editing levels similar to the on-targets GAPDH (41%) and ACTB (52%), see Supplementary Table S7. This was due to the high sequence homology between ACTA2 and ACTB. In order to spot even minute guideRNA-dependent bystander editings, we manually analyzed the regions around the two on-target sites (± 500 bp) without applying the usual cutoff for editing difference. This yielded 4 bystander sites in GAPDH (editing difference $\leq$ 1%) and 10 sites in ACTB, with the three highest sites exhibiting editing differences between 16.0% and 7.7%, likely due to high similarity with the on-target site (Supplementary Tables S8 and S9, Supplementary Figure S20). Overall, NGS analysis demonstrated again (9,10) that total off-target effects are dominated by guideRNA-

Table 1. Number of significantly differently edited sites found in editing experiments in mono cell lines SA1Q, SA2Q, and in duo cell line **2** (HA1Q + SA2Q) in comparison to a negative control cell line (293 Flp-In T-REx) not expressing any editing enzyme (Total off-targets). The last column shows the guideRNA-dependent fraction of the total off-targets for duo cell line **2**

	Total off-targets			gRNA-dependent
	SA1Q	SA2Q	HA1Q + SA2Q	HA1Q + SA2Q
Total number	3406	4795	8391	653
incl. Alu sites	400	1190	1281	136
5′UTR	124	168	286	19
Missense mutation	769	1080	2150	166
Nonstop mutation	51	46	108	5
Start codon SNP	1	1	2	0
Silent	470	515	1079	74
3′UTR	1427	2009	3422	267
Noncoding	564	976	1343	122

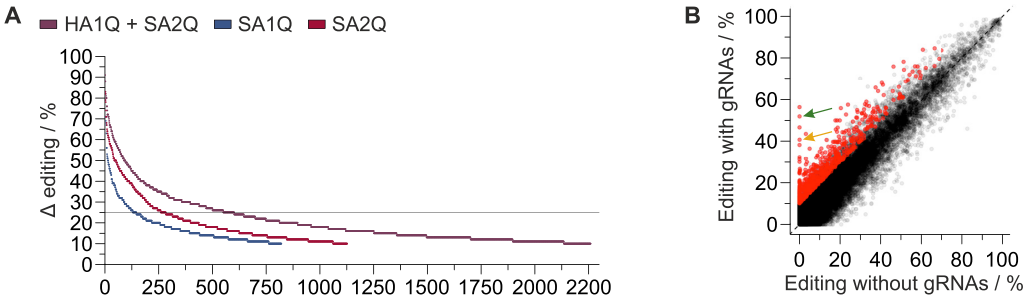


Figure 4. Off-target analysis of duo cell line **2**. (**A**) Total off-target editing of duo cell line **2** (HA1Q + SA2Q) in comparison with mono cell lines SA1Q and SA2Q. Shown are significantly differently edited sites (≥ 10% editing difference, Fisher's exact test, two-sided, adjusted $P < 0.01$, $n = 2$ experiments) that led to nonsynonymous substitutions, sorted by editing difference. (**B**) Scatter plot depicting the guideRNA-dependent off-target effects in duo cell line **2**. Significantly differently edited sites are marked in red. The two on-target sites (in ACTB and GAPDH) are marked by a green and yellow arrow respectively.

independent off-target effects rather than by mis-guiding through the guideRNAs.

Selective site-directed C-to-U and A-to-I editing can be combined within one cell

C-to-U and A-to-I RNA base editing complement one another. While A-to-I editing can remove premature STOP codons, C-to-U editing can write them and furthermore affect different amino acid substitutions, including key residues like serine and proline. APOBEC1-mediated C-to-U RNA editing plays a key role for human physiology by inducing an isoform switch in ApoB48/100 (37). In preliminary experiments, we found that a simple fusion of the SNAP-tag to the C-terminus of murine APOBEC1 generates an effector protein dubbed APO1S that can induce C-to-U editing in an RNA-guided manner. Fully analog to the duo cell lines above, we generated four duo cell lines (**6**–**9**, Figure 5A) that co-express the HA1Q and APO1S transgenes under control of doxycycline. Via western blot/SDS PAGE we characterized the relative transgene expression (Figure 5B), which suggested cell line **6** and **9** to express sufficient levels of both effectors. Notably, the inserts of cell lines **6** and **9** are constructed analog to those in cell lines **2** and **5**, indicating that these two designs might be generally applicable for the co-expression of two RNA-guided effector proteins.

A first set of editing experiments targeted a 5′-UAG codon for HA1Q-mediated A-to-I editing and a proximal 5′-ACG codon for APO1S-mediated C-to-U editing in an eGFP reporter transcript in duo cell line **9**. The target sites are close enough to design one guideRNA that can mediate both, adenosine or cytidine deamination, depending on the self-labeling moiety attached, since HA1Q requires an RNA duplex as substrate (9) whereas APO1S prefers its positioning 4–6 nt upstream of the target site (Figure 5C). As expected, the halo-eGFP guideRNA elicited A-to-I editing, the snap-eGFP guideRNA elicited C-to-U editing and a bisfunctional halo-snap-eGFP guideRNA induced both A-to-I and C-to-U editing (Figure 5D). Similar results have been obtained in the cell lines **6** and **7** (Supplementary Figure S14). Notably, the snap-eGFP guideRNA also induced some A-to-I editing. However, highly selective C-to-U editing was achieved when a snap-eGFP guideRNA was applied that was fully chemically modified (*mod*-snap-eGFP, Figure 5C, Supplementary Table S1) and that did not contain the modification gap (38) around the adenosine required for ADAR1 action (Figure 5D). This highlights another strength of the RNA targeting platform. Bystander off-target editing can be easily controlled by chemical modification of the guideRNA (9), a frequent problem (8,10) with RNA base editing approaches that apply genetically encoded guideRNAs.

In a second set of editing experiments, we applied two different guideRNAs to selectively recruit APO1S and HA1Q to two different endogenous transcripts in duo cell line **9**. The $(halo)_2$-UAU guideRNA steers HA1Q to edit the adenosine in a 5′-UAU codon in the ORF of ACTB, the $(snap)_2$-ACC guideRNA steers APO1S to edit the cytosine in a 5′-ACC codon in the ORF of GAPDH (Figure 5E). In contrast to the editing of the eGFP reporter, editing on endogenous ORF targets was very selective. The $(halo)_2$-UAU guideRNA induced site-specific A-to-I editing with excellent yields (ca. 65%) in the ACTB transcript with no detectable C-to-U editing, whereas the $(snap)_2$-ACC guideRNA induced site-specific C-to-U editing with moderate yield (ca. 20%) in the GAPDH transcript, again with no detectable A-to-I RNA editing (Figure 5F). Notably, co-transfection of both guideRNAs induced selective A-to-I and C-to-U editing in the ACTB and GAPDH transcript, respectively, without any loss of editing efficiency compared to the single guideRNA transfections. Similar results have been obtained in the cell lines **6** and **7** (Supplementary Figure S13e). Thus, concurrent C-to-U and A-to-I editing can be done within one cell under programmable target selection.

We then benchmarked the C-to-U editing efficiency achieved with APO1S in duo cell line **9** at both targets (eGFP and GAPDH) with the recently published (39) Cas13-based RESCUE approach (Supplementary Figure S15). Specifically, we tested the most active variant, RESCUEr16, and tried four different C-flip guideRNAs for each target (Supplementary Table S4). The APO1S enzyme outcompeted RESCUEr16 on both targets with respect to on-target editing yield. While we found C-to-U bystander editing for both approaches, only the RESCUE approach induced A-to-I bystander editing (Supplementary Table S5).

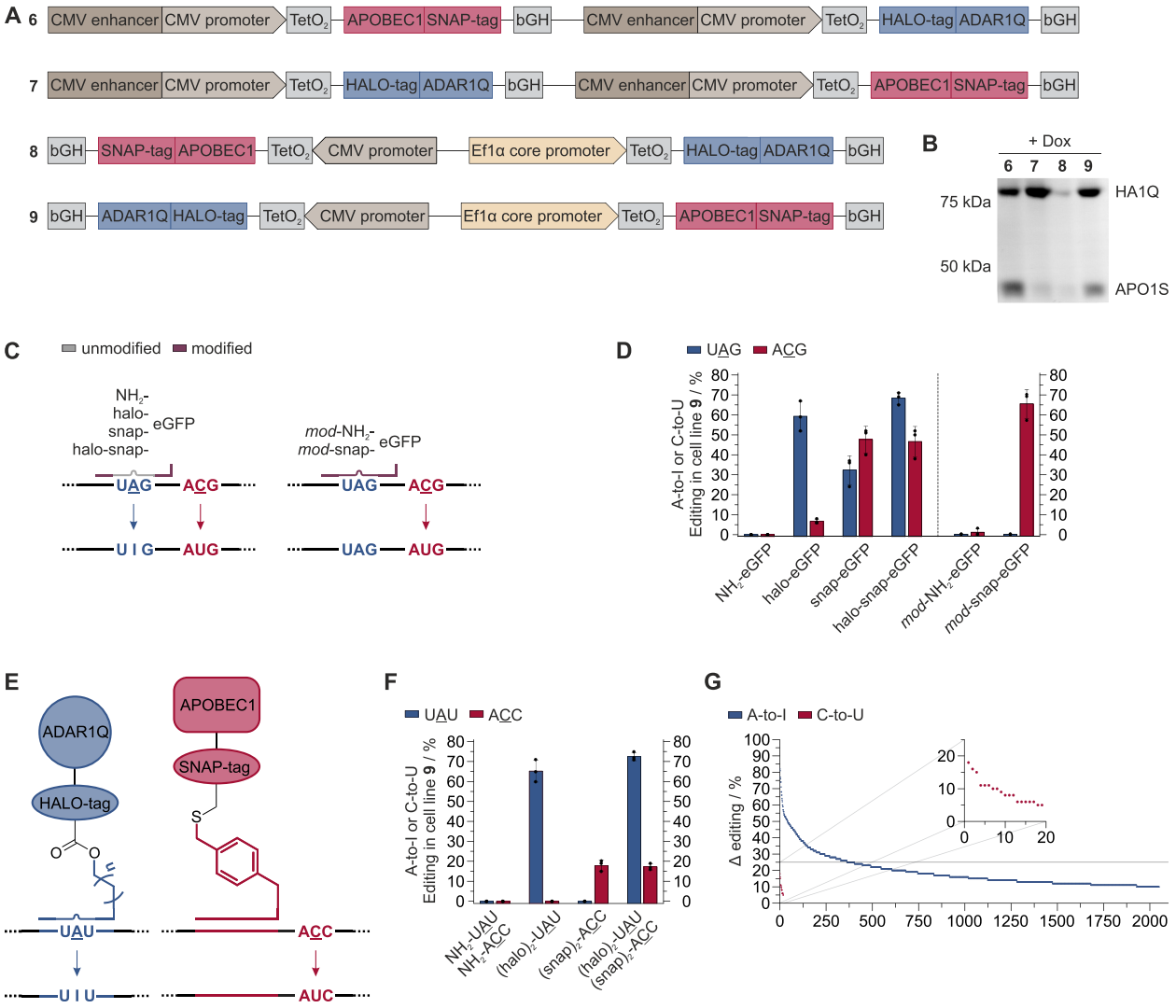


Figure 5. Selective and concurrent A-to-I and C-to-U editing. (**A**) Constructs (**6**-**9**) were designed to co-express both transgenes (APO1S and HA1Q) from one cassette under doxycycline control. $TetO_2$: tet operator, leads to repression of expression in the absence of a tetracycline (33); bGH: bovine growth hormone terminator; P2A: porcine teschovirus-1 self-cleaving 2A peptide (32). (**B**) Characterization of relative transgene expression via SDS-PAGE after co-staining with TMR-BG and TMR-chloroalkane in raw cell lysate. (**C**) GuideRNA design to enable or block concurrent A-to-I and C-to-U editing in an eGFP reporter with a single guideRNA. The modified guideRNA (*mod*-snap-eGFP) contained chemical modifications (Supplementary Table S1) that block A-to-I editing. (**D**) Editing yield in cell line **9** from concurrent A-to-I and C-to-U editing in an eGFP reporter transcript after transfection of a halo-, snap- or halo-snap-guideRNA (5.0 pmol). (**E**) GuideRNA design and recruiting strategy for concurrent and selective A-to-I and C-to-U editing at two different endogenous transcripts. (**F**) Editing yield in cell line **9** for selective and concurrent editing as depicted in E) after transfection of a single or co-transfection of two guideRNAs, one $(halo)_2$-guideRNA for A-to-I editing in ACTB and one $(snap)_2$-guideRNA for C-to-U editing in GAPDH (5.0 pmol each). Data in D) and F) are shown as the mean ± SD of $N = 3$ independent experiments. (**G**) Total off-target A-to-I and C-to-U editing of duo cell line **9**. Shown are significantly differently edited sites (for A-to-I ≥ 10% editing difference, for C-to-U ≥ 5% editing difference, Fisher's exact test, two-sided, adjusted $P < 0.01$, $n = 2$ experiments) that led to nonsynonymous substitutions, sorted by editing difference. A-to-I (in ACTB) and C-to-U (in GAPDH) on-target sites are no. 40 and no. 18, respectively.

To assess transcriptome-wide global A-to-I and C-to-U off-target editing, we applied next generation RNA sequencing to detect significantly differently edited sites in duo cell line **9** after co-transfection of 2.5 pmol $(halo)_2$-UAU and 2.5 pmol $(snap)_2$-ACC guideRNA in comparison to a cell line lacking expression of any artificial editing enzyme (Table 2, Figure 5g, Supplementary Figure S22). Expressing only one A-to-I editing enzyme (HA1Q), the total number of A-to-I off-target sites (6767) was below that of duo cell line **2**, which expresses two A-to-I editing enzymes. Again, the majority of sites exhibited differences in editing below 25% (Supplementary Table S10). A slightly higher fraction of the off-target sites was guideRNA-dependent compared to cell line **2**, which might be due to the higher guideRNA amounts applied in cell line **9**. However, in particular off-target sites with high editing differences, e.g. ≥ 25%, were typically guideRNA-independent (Supplementary Table S10). Taking the generally lower C-to-U editing yields into account, we adapted the pipeline and set the cutoff for editing differences to 5%. With this highly sen-

Table 2. Number of significantly differently edited A-to-I and C-to-U sites found in editing experiments in duo cell line **9** (HA1Q + APO1S) in comparison to a negative control cell line (293 Flp-In T-REx) not expressing any editing enzyme (Total off-targets). The guideRNA-dependent fractions of the total off-targets are shown in the right column, respectively

	A-to-I ($\Delta \geq 10\%$)		C-to-U ($\Delta \geq 5\%$)	
	Total off-targets	gRNA-dependent	Total off-targets	gRNA-dependent
Total number	6767	2148	2976	153
incl. Alu sites	729	85	17	1
5′UTR	262	92	44	3
Missense mutation	1944	704	16	1
Nonsense mutation	0	0	2	0
Nonstop mutation	104	30	0	0
Start codon SNP	2	0	0	0
Silent	979	352	17	2
3′UTR	2560	731	2593	131
Noncoding	916	239	304	16

sitive pipeline, we were able to find 2976 significantly differently edited sites (Table 2). However, the vast number of sites showed editing differences below 10%, and only 129 sites had editing differences above 25% (Supplementary Table S11). Notably, almost all off-target sites were located in the 3′-UTR, and only 18 of 2976 total sites were inducing missense or nonsense mutations. Also the number of guideRNA-dependent off-targets sites was comparably low (153 of 2976), with basically all in the 3′-UTR (Table 2, Supplementary Table S11). Again, we manually analyzed the regions (± 500 bp) around the on-target site to detect low-level bystander A-to-I editing in ACTB (Supplementary Table S13) and C-to-U bystander editing in GAPDH (Supplementary Table S12). We found one bystander site in ACTB (editing difference $\leq 1\%$) and a larger number (22) of bystander sites in GAPDH, but only one of the 22 sites had an editing difference $\geq 1\%$. Overall, our approach for concurrent A-to-I and C-to-U RNA editing, based on co-expression of two different editing enzymes gave moderate, mainly guideRNA-independent off-target effects for both effectors.

CONCLUSIONS

Here, we show for the first time that one can combine two self-labeling enzymes to create a powerful RNA targeting platform to manipulate RNA inside living cells in a yet unprecedented way. The orthogonality of HALO- and SNAP-tag sets the ground for the selective and programmable steering of two different RNA effectors. Furthermore, the approach benefits from the small size of the fusion proteins, which enable their facile genomic co-integration, and the ease by which the short (20 nt), chemically stabilized guideRNAs can be co-transfected and optimized to reduce bystander editing, if required. Recent attempts to combine two base editing activities in one protein either to target DNA (40) or RNA (39) illustrate the manifold problems of controlling two enzyme functions independently, which we could solve here for RNA base editing. We successfully demonstrate the functioning of our approach for the orthogonal and concurrent recruitment of two pairs of editing effectors. The selective recruitment of ADAR1 and ADAR2 deamination activity enables site-directed A-to-I RNA base editing with improved editing efficiency. The selective recruitment of ADAR1 and APOBEC1 deamination activity allows for target-selective, concurrent A-to-I and C-to-U editing. Notably, orthogonality is particularly effective with guideRNAs that can recruit two copies of an editase. Again, we demonstrate that genetic integration of the editing enzymes helps to control global off-target A-to-I and C-to-U editing induced by unengaged editing enzymes (9,10,16,41). Notably, even the concurrent transfection of two guideRNAs leads to only a very small number of off-target editing events caused by misguiding through the guideRNAs, and might be amenable for further sequence optimization, if required.

Furthermore, our platform benefits from the high flexibility in the linker chemistry. This makes it possible to control the composition and stoichiometry of two fusion proteins at a target with one guideRNA. We exemplify this with the generation of bisfunctional guideRNAs that are capable of co-recruiting either ADAR1/ADAR2 or ADAR1/APOBEC1 to one target with one guideRNA. The possibility of including photochemistry to the linker may add another level of spatio-temporal control in the future (18,19). The general concept we present here may be readily transferred to recruit further pairs of writers and erasers of epitranscriptomic marks with ease and unprecedented control (2,42,43).

DATA AVAILABILITY

NGS raw data for duo cell lines **2** and **9** can be found on NCBI GEO under GSE160945**.** The online Supplementary Information contains an excel sheet with all editing yields, an excel sheet with the NGS analysis, SnapGene files for all transgenic constructs, and a PDF file giving detailed information on syntheses, protocols, and additional experimental data.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) [430214260, STA1053/7-1 to T.S.]; DFG priority program SPP 1784 [404867268 to F.N.P., T.S.]; European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program [649019 to F.N.P., 647328 to T.S.]. Funding for open access charge: ERC; DFG.

Conflict of interest statement. T.S. holds patents on site-directed RNA editing.

REFERENCES

1. Corbett,A.H. (2018) Post-transcriptional regulation of gene expression and human disease. *Curr. Opin. Cell Biol.*, **52**, 96–104.
2. Kadumuri,R.V. and Janga,S.C. (2018) Epitranscriptomic code and its alterations in human disease. *Trends Mol. Med.*, **24**, 886–903.
3. Gagnidze,K., Rayon-Estrada,V., Harroch,S., Bulloch,K. and Papavasiliou,F.N. (2018) A new chapter in genetic medicine: RNA

editing and its role in disease pathogenesis. *Trends Mol. Med.*, **24**, 294–303.
4. Shi,H., Zhang,X., Weng,Y.-L., Lu,Z., Liu,Y., Lu,Z., Li,J., Hao,P., Zhang,Y., Zhang,F. *et al.* (2018) m^{6}A facilitates hippocampus-dependent learning and memory through YTHDF1. *Nature*, **563**, 249–253.
5. Frye,M., Harada,B.T., Behm,M. and He,C. (2018) RNA modifications modulate gene expression during development. *Science*, **361**, 1346–1349.
6. Liu,J., Eckert,M.A., Harada,B.T., Liu,S.-M., Lu,Z., Yu,K., Tienda,S.M., Chryplewicz,A., Zhu,A.C., Yang,Y. *et al.* (2018) m^{6}A mRNA methylation regulates AKT activity to promote the proliferation and tumorigenicity of endometrial cancer. *Nat. Cell Biol.*, **20**, 2085–1083.
7. Ishizuka,J.J., Manguso,R.T., Cheruiyot,C.K., Bi,K., Panda,A., Iracheta-Vellve,A., Miller,B.C., Du,P.P., Yates,K.B., Dubrot,J. *et al.* (2019) Loss of ADAR1 in tumours overcomes resistance to immune checkpoint blockade. *Nature*, **565**, 43–48.
8. Rees,H.A. and Liu,D.R. (2018) Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.*, **19**, 770–788.
9. Vogel,P., Moschref,M., Li,Q., Merkle,T., Selvasaravanan,K.D., Li,J.B. and Stafforst,T. (2018) Efficient and precise editing of endogenous transcripts with SNAP-tagged ADARs. *Nat. Methods*, **15**, 535–538.
10. Vogel,P. and Stafforst,T. (2019) Critical review on engineering deaminases for site-directed RNA editing. *Curr. Opin. Biotechnol.*, **55**, 74–80.
11. Kuttan,A. and Bass,B.L. (2012) Mechanistic insights into editing-site specificity of ADARs. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, E3295–E3304.
12. Keppler,A., Gendreizig,S., Gronemeyer,T., Pick,H., Vogel,H. and Johnsson,K. (2003) A general method for the covalent labeling of fusion proteins with small molecules *in vivo*. *Nat. Biotechnol.*, **21**, 86–89.
13. Stafforst,T. and Schneider,M.F. (2012) An RNA–Deaminase conjugate selectively repairs point mutations. *Angew. Chem.*, **124**, 11329–11332.
14. Abudayyeh,O.O., Gootenberg,J.S., Konermann,S., Joung,J., Slaymaker,I.M., Cox,D.B.T., Shmakov,S., Makarova,K.S., Semenova,E., Minakhin,L. *et al.* (2016) C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science*, **353**, 5573.
15. Abudayyeh,O.O., Gootenberg,J.S., Essletzbichler,P., Han,S., Joung,J., Belanto,J.J., Verdine,V., Cox,D.B.T., Kellner,M.J., Regev,A. *et al.* (2017) RNA targeting with CRISPR–Cas13. *Nature*, **550**, 280–284.
16. Cox,D.B.T., Gootenberg,J.S., Abudayyeh,O.O., Franklin,B., Kellner,M.J., Joung,J. and Zhang,F. (2017) RNA editing with CRISPR-Cas13. *Science*, **358**, 1019–1027.
17. Montiel-Gonzalez,M.F., Vallecillo-Viejo,I., Yudowski,G.A. and Rosenthal,J.J.C. (2013) Correction of mutations within the cystic fibrosis transmembrane conductance regulator by site-directed RNA editing. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 18285–18290.
18. Hanswillemenke,A., Kuzdere,T., Vogel,P., Jékely,G. and Stafforst,T. (2015) 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.
19. Vogel,P., Hanswillemenke,A. and Stafforst,T. (2017) Switching protein localization by site-directed RNA editing under control of light. *ACS Synth. Biol.*, **6**, 1642–1649.
20. Vogel,P. and Stafforst,T. (2014) Site-directed RNA editing with antagomir deaminases — a tool to study protein and RNA function. *ChemMedChem*, **9**, 2021–2025.
21. Hanswillemenke,A. and Stafforst,T. (2019) Protocols for the generation of caged guideRNAs for light-triggered RNA-targeting with SNAP-ADARs. *Methods Enzymol.*, **624**, 47–68.
22. Dobin,A., Davis,C.A., Schlesinger,F., Drenkow,J., Zaleski,C., Jha,S., Batut,P., Chaisson,M. and Gingeras,T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.
23. Kent,W.J., Sugnet,C.W., Furey,T.S., Roskin,K.M., Pringle,T.H., Zahler,A.M. and Haussler,D. (2002) The human genome browser at UCSC. *Genome Res.*, **12**, 996–1006.
24. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G. and Durbin,R. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
25. Lo Giudice,C., Tangaro,M.A., Pesole,G. and Picardi,E. (2020) Investigating RNA editing in deep transcriptome datasets with REDItools and REDIportal. *Nat. Protoc.*, **15**, 1098–1131.
26. Picardi,E. and Pesole,G. (2013) REDItools: high-throughput RNA editing detection made easy. *Bioinformatics*, **29**, 1813–1814.
27. Ramos,A.H., Lichtenstein,L., Gupta,M., Lawrence,M.S., Pugh,T.J., Saksena,G., Meyerson,M. and Getz,G. (2015) Oncotator: cancer variant annotation tool. *Hum. Mutat.*, **36**, E2423–E2429.
28. Los,G.V., Encell,L.P., McDougall,M.G., Hartzell,D.D., Karassina,N., Zimprich,C., Wood,M.G., Learish,R., Ohana,R.F., Urh,M. *et al.* (2008) HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS Chem. Biol.*, **3**, 373–382.
29. Gautier,A., Juillerat,A., Heinis,C., Corrêa,I.R. Jr, Kindermann,M., Beaufils,F. and Johnsson,K. (2008) An engineered protein tag for multiprotein labeling in living cells. *Chem. Biol.*, **15**, 128–136.
30. Eggington,J.M., Greene,T. and Bass,B.L. (2011) Predicting sites of ADAR editing in double-stranded RNA. *Nat. Commun.*, **2**, 319–319.
31. Palmer,A.C., Egan,J.B. and Shearwin,K.E. (2011) Transcriptional interference by RNA polymerase pausing and dislodgement of transcription factors. *Transcription*, **2**, 9–14.
32. Kim,J.H., Lee,S.-R., Li,L.-H., Park,H.-J., Park,J.-H., Lee,K.Y., Kim,M.-K., Shin,B.A. and Choi,S.-Y. (2011) High cleavage efficiency of a 2A peptide derived from porcine Teschovirus-1 in human cell lines, Zebrafish and Mice. *PLoS One*, **6**, e18556.
33. Yao,F., Svensjö,T., Winkler,T., Lu,M., Eriksson,C. and Eriksson,E. (1998) Tetracycline repressor, tetR, rather than the tetR–mammalian cell transcription factor fusion derivatives, regulates inducible gene expression in mammalian cells. *Hum. Gene Ther.*, **9**, 1939–1950.
34. Nishikura,K. (2010) Functions and regulation of RNA editing by ADAR deaminases. *Annu. Rev. Biochem.*, **79**, 321–349.
35. Tan,M.H., Li,Q., Shanmugam,R., Piskol,R., Kohler,J., Young,A.N., Liu,K.I., Zhang,R., Ramaswami,G., Ariyoshi,K. *et al.* (2017) Dynamic landscape and regulation of RNA editing in mammals. *Nature*, **550**, 249–254.
36. Montiel-González,M.F., Vallecillo-Viejo,I.C. and Rosenthal,Joshua J.C. (2016) An efficient system for selectively altering genetic information within mRNAs. *Nucleic Acids Res.*, **44**, e157.
37. Gott,J.M. and Emeson,R.B. (2000) Functions and mechanisms of RNA editing. *Annu. Rev. Genet.*, **34**, 499–531.
38. Vogel,P., Schneider,M.F., Wettengel,J. and Stafforst,T. (2014) Improving site-directed RNA editing in vitro and in cell culture by chemical modification of the GuideRNA. *Angew. Chem.*, **126**, 6382–6386.
39. Abudayyeh,O.O., Gootenberg,J.S., Franklin,B., Koob,J., Kellner,M.J., Ladha,A., Joung,J., Kirchgatterer,P., Cox,D.B.T. and Zhang,F. (2019) A cytosine deaminase for programmable single-base RNA editing. *Science*, **365**, 382–386.
40. Sakata,R.C., Ishiguro,S., Mori,H., Tanaka,M., Tatsuno,K., Ueda,H., Yamamoto,S., Seki,M., Masuyama,N., Nishida,K. *et al.* (2020) Base editors for simultaneous introduction of C-to-T and A-to-G mutations. *Nat. Biotechnol.*, **38**, 865–869.
41. Vallecillo-Viejo,I.C., Liscovitch-Brauer,N., Montiel-Gonzalez,M.F., Eisenberg,E. and Rosenthal,J.J.C. (2018) Abundant off-target edits from site-directed RNA editing can be reduced by nuclear localization of the editing enzyme. *RNA Biol.*, **15**, 104–114.
42. Rau,K., Rösner,L. and Rentmeister,A. (2019) Sequence-specific m^{6}A demethylation in RNA by FTO fused to RCas9. *RNA*, **25**, 1311–1323.
43. Liu,X.-M., Zhou,J., Mao,Y., Ji,Q. and Qian,S.-B. (2019) Programmable RNA N^6-methyladenosine editing by CRISPR-Cas9 conjugates. *Nat. Chem. Biol.*, **15**, 865–871.

Supporting Information of Manuscript 2

Harnessing self-labeling enzymes for selective and concurrent A-to-I and C-ot-U RNA base editing

Anna S. Stroppel, **Ngadhnjim Latifi**, Alfred Hanswillemenke, Rafail Nikolaos Tasakis, F. Nina Papavasiliou and Thorsten Stafforst, *Nucleic Acids Research*, 2021, *49.16,* e95

Supporting Information

Harnessing Self-Labeling Enzymes for Selective and Concurrent A-to-I and C-to-U RNA Base Editing

Anna S. Stroppel[1], Ngadhnjim Latifi[1], Alfred Hanswillemenke[1], Rafail Nikolaos Tasakis[2,3], F. Nina Papavasiliou[2,3] and Thorsten Stafforst[1*]

[1]Interfaculty Institute of Biochemistry, University of Tübingen, Auf der Morgenstelle 15, 72076 Tübingen, Germany

[2]Division of Immune Diversity (D150), German Cancer Research Center (DKFZ), Heidelberg, Germany.

[3]Faculty of Biosciences, University of Heidelberg, Heidelberg, Germany

[*]correspondence to thorsten.stafforst@uni-tuebingen.de

Contents

List of Schemes

List of Figures

List of Tables

Chemical synthesis

General

All chemicals were purchased from standard chemical providers and used without further purification unless stated otherwise. Reactions that are sensible towards air or water were carried out with anhydrous solvents and under a nitrogen atmosphere using Schlenk technique.

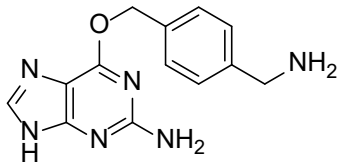
For TLC, silica gel F_{254} foils from MERCK were used, which were visualized either under UV light at 254 nm or with 0.5 % aqueous solution of $KMnO_4$ or with 0.1 % aqueous solution of ninhydrin supplemented with 10 % ethanol. Purification by column chromatography was performed with self-packed columns of silica gel (0.04 – 0.063 mm/230 – 240 mesh), applying slight overpressure.

Analytical as well as preparative HPLC was conducted with a system by SHIMADZU consisting of a SCL-10A VP system controller, two LC-20AT prominence liquid chromatographs for buffers A and B and a SPD-20AV prominence UV/VIS detector. Buffer A consisted of H_2O:TFA, 100:0.1, buffer B of MeCN:H_2O:TFA, 90:10:0.1. For analytic measurements, a linear gradient from 5 % B to 95 % B in 25 min was applied. As analytical column, an EC 125/4 nucleodur 100-5 C18 ec column by MACHEREY-NAGEL was used, as preparative a VP 250/10 nucleodur 100-5 C18 ec column by MACHEREY-NAGEL. Spectra were analyzed with SHIMADZU CLASS-VP.

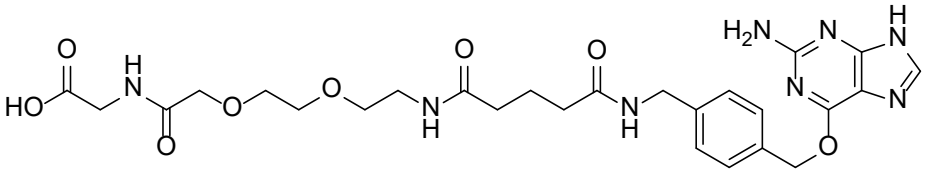
NMR spectra were measured on a BRUKER Avance III HD 300 spectrometer at 300.13 MHz or a BRUKER Avance III HDX 400 spectrometer at 400.16 MHz for ^{1}H spectra or 100.62 MHz for ^{13}C spectra respectively. Chemical shifts in ppm were calibrated to the signal of the deuterated solvent. Melting points were determined with a Melting Point M-560 from BÜCHI. UV spectra were measured with a Cary 300 Scan UV/Visible spectrophotometer from Agilent.

LC/MS spectra were recorded on a SHIMADZU LCMS-2020 with kinetex C18 column. Buffer A consisted of H_2O:HCO_2H, 100:0.1, buffer B of MeCN:H_2O:HCO_2H, 80:20:0.1. A linear gradient from 5 % B to 95 % B in 10 min was applied. For high resolution HR-ESI-TOF mass spectra, a BRUKER Daltonics maxis 4G mass spectrometer was used. Elemental analysis was performed with the elemental analyser Euro EA 3000 from HEKATECH.

BG-linker-OH (snap)



BG-NH$_2$

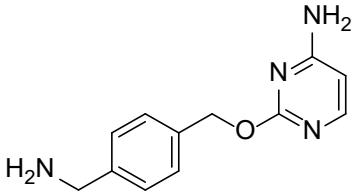


BG-linker-OH

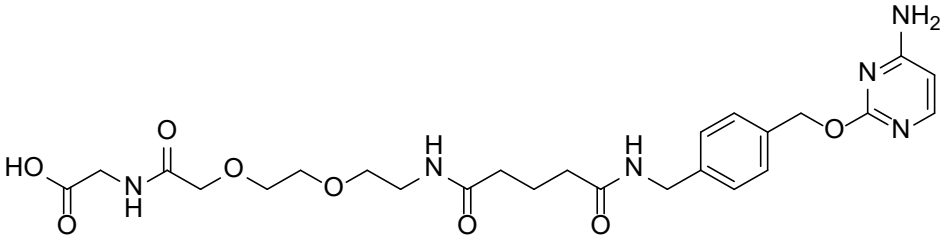
Scheme S1. Structures of BG-NH2 and BG-linker-OH (snap).

Literature known O^6-(4-aminomethyl-benzyl)guanine (BG-NH$_2$)[1] and BG-linker-OH (snap)[2] were synthesized starting from commercially available 6-chloro-guanine according to previously reported protocols.

BC-linker-OH (clip)



BC-NH$_2$



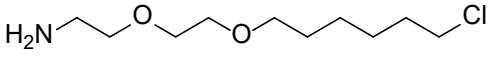
BC-linker-OH

Scheme S2. Structures of BC-NH$_2$ and BC-linker-OH (clip).

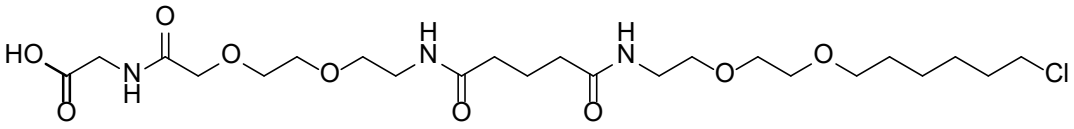
2-(4-(Aminomethyl)-benzyloxy)-4-aminopyrimidine (BC-NH_2) was prepared from commercially available methyl-4-(aminomethyl)benzoate hydrochloride according to literature.[3-5] BC-linker-OH (clip) was obtained via the solid phase peptide synthesis protocol described for BG-linker-OH[2] starting from 413 mg (260 µmol, 1.00 eq) H-Gly-2-chlorotrityl resin by using 20 mg (87.0 µmol, 0.33 eq) BC-NH_2 in step 7. The resulting crude product was dissolved in 20 % buffer B, filtered and purified via preparative HPLC (5 % B to 40 % B in 40 min), which yielded 31 mg (58.8 µmol, 65 %) BC-linker-OH as a colorless powder after lyophilization.

mp = 111.4 °C (H_2O/MeCN); UV spectrum (MeOH): λ_{max} = 270 nm, $\varepsilon_{260\,nm}$ = 4.24 $mM^{-1}cm^{-1}$; t_R = 6.8 min; *m/z* calculated for $[C_{25}H_{34}N_6O_8+H]^+$: 547.25109, found: 547.25159.

Chloroalkane-linker-OH (halo)



chloroalkane-NH_2



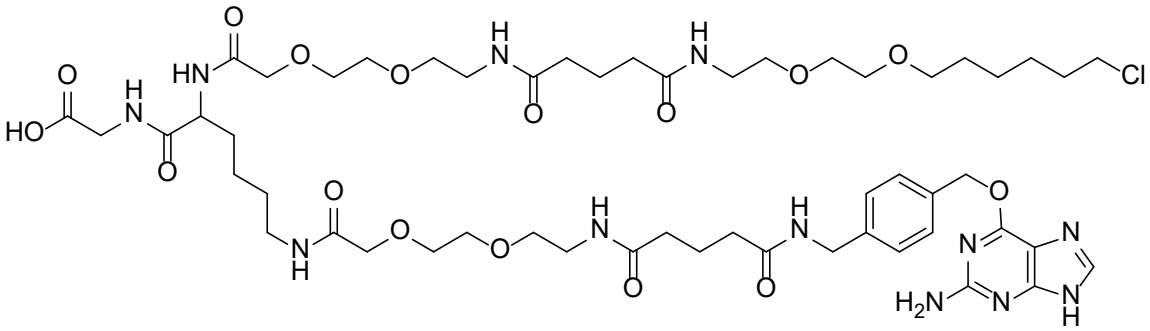
chloroalkane-linker-OH

Scheme S3. Structures of chloroalkane-NH_2 and chloroalkane-linker-OH (halo).

2-(2-((6-Chlorohexyl)oxy)ethoxy)ethanamine (chloroalkane-NH_2) was synthesized from commercially available 2-(2-aminoethoxy)ethanol according to literature.[6] Chloroalkane-linker-OH (halo) was obtained via the protocol described for BG-linker-OH[2] starting from 413 mg (260 µmol, 1.00 eq) H-Gly-2-chlorotrityl resin by using 19 mg (85.8 µmol, 0.33 eq) chloroalkane-NH_2 in step 7. The resulting crude product was dissolved in 10 % buffer B, filtered and purified via preparative HPLC (15 % B to 70 % B in 40 min), which yielded 12 mg (22.0 µmol, 26 %) chloroalkane-linker-OH as a colorless powder after lyophilization.

mp = 67.2 °C (H_2O/MeCN); t_R = 11.7 min; *m/z* calculated for $[C_{23}H_{42}ClN_3O_9+Na]^+$: 562.25018, found: 562.25036. Elemental analysis: calculated: C: 51.15 %, H: 7.84 %, N: 7.78 %, found: C: 51.11 %, H: 7.95 %, N: 7.87 %.

Chloroalkane-BG-linker-OH (halo-snap)



chloroalkane-BG-linker-OH

Figure S1. Structure of chloroalkane-BG-linker-OH (halo-snap).

All reaction steps were performed at room temperature on a peptide shaker at 1000 rpm. Unless indicated otherwise, washing refers to washing 3 × each with 5 ml NMP/DCM (1:1), then DCM, then NMP.

In a 10 ml syringe with a polyethylene frit, 75 mg (40.6 µmol, 1.00 eq) H-Gly-2-chlorotrityl resin were swelled in NMP for 1 h. 92 mg (203 µmol, 5.00 eq) FmocLys(Alloc)-OH, 69 mg (183 µmol, 4.50 eq) HBTU, 27 mg (203 µmol, 5.00 eq) HOBt and 240 µl (184 mg, 1.42 mmol, 35.0 eq) DIPEA in 1.5 ml NMP were added to the resin and shaken for 1 h. After washing, 3 × 6 ml 20 % piperidine in NMP were applied for 10 min each for Fmoc deprotection and the resin was washed again. Then, 63 mg (162 µmol, 4.00 eq) Fmoc-AEEA-COOH was coupled to the resin with 55 mg (146 µmol, 3.60 eq) HBTU, 22 mg (162 µmol, 4.00 eq) HOBt and 193 µl (147 mg, 1.14 mmol, 28.0 eq) DIPEA in 1 ml NMP for 1 h. After washing, 3 × 6 ml 20 % piperidine in NMP were applied for 10 min each for Fmoc deprotection and the resin was washed again. 59 mg (406 µmol, 10.0 eq) glutaric anhydride was coupled with 69 µl (52 mg, 406 µmol, 10.0 eq) DIPEA in 1 ml NMP for 30 min. The resin was washed again with a subsequent additional washing step with 1 % NaOH in dioxane/H_2O (1:1) for 1 min and thereafter washing 4 × each with DCM/NMP (1:1), DCM and NMP. For activation of the glutaric acid, 2 × 73 µl (119 mg, 426 µmol, 10.5 eq) Pfp-OTFA in 2 ml pyridine/DCM (1:1) were applied for 10 min and 20 min respectively. After washing 4 × with NMP only, 25 mg (112 µmol, 2.75 eq) chloroalkane-NH_2 with 250 µl (190 mg, 1.47 mmol, 36.2 eq) in 1 ml NMP were coupled to the resin overnight.

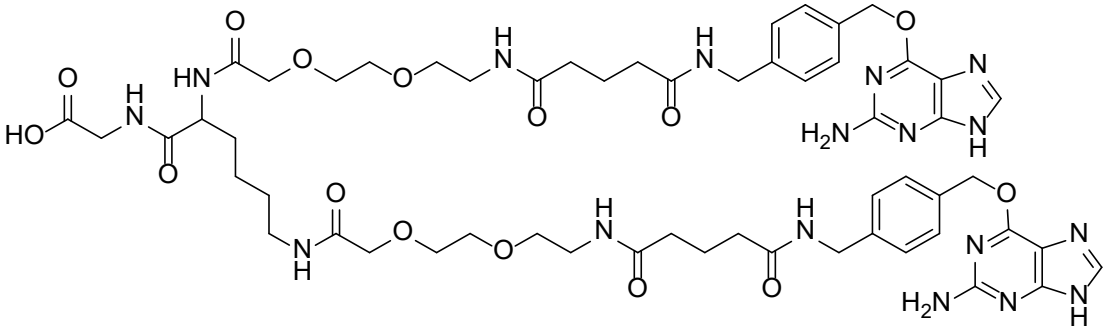
For Alloc deprotection, the resin was washed again, rendered inert under nitrogen atmosphere and additionally washed 5 × with anhydrous DCM for 30 s. First, 119 µl (105 mg, 975 µmol, 24.0 eq) $PhSiH_3$ in 1 ml anhydrous DCM were applied to the resin, then 4.7 mg (4.06 µmol, 0.10 eq) $Pd(PPh_3)_4$ in 1.5 ml NMP were added. After 10 min, the resin was washed 8 × with anhydrous DCM for 30 s and the procedure was repeated once.

63 mg (162 µmol, 4.00 eq) Fmoc-AEEA-COOH was coupled to the resin with 55 mg (146 µmol, 3.60 eq) HBTU, 22 mg (162 µmol, 4.00 eq) HOBt and 193 µl (147 mg, 1.14 mmol, 28.0 eq) DIPEA in 1 ml NMP for 1 h. After washing, 3 × 6 ml 20 % piperidine in NMP were applied for 10 min each for Fmoc deprotection and the resin was washed again. 59 mg (406 µmol, 10.0 eq) glutaric anhydride was coupled with 69 µl (52 mg, 406 µmol, 10.0 eq) DIPEA in 1 ml NMP for 30 min. The resin was washed again with a subsequent additional washing step with 1 % NaOH in dioxane/H_2O (1:1) for 1 min and thereafter washing 4 × each with DCM/NMP (1:1), DCM and NMP. For activation of the glutaric acid, 2 × 73 µl (119 mg, 426 µmol, 10.5 eq) Pfp-OTFA in 2 ml pyridine/DCM (1:1) were applied for 10 min and 20 min respectively. After washing 4 × with NMP only, 30 mg (112 µmol, 2.75 eq) BG-NH_2 in 1 ml DMSO/pyridine (20:1) were coupled to the resin overnight.

Finally, the resin was washed 4 × each with NMP/DCM (1:1), DCM, NMP and Et_2O, swelled 2 × in NMP for 5 min and washed 4 × with DCM. Cleavage from the resin was executed under continuous flow with 20 ml DCM/HFIP (8:2), which was removed under reduced pressure. The resulting crude product was dissolved in 30 % buffer B, filtered and purified via preparative HPLC (30 % B to 65 % B in 55 min), which yielded 15 mg (12.7 µmol, 31 %) chloroalkane-BG-linker-OH as a colorless powder after lyophilisation.

mp = 93.2 °C (H_2O); UV spectrum (PBS): λ_{max} = 283 nm, $\varepsilon_{260\ nm}$ = 2.50 $mM^{-1}cm^{-}1$; t_R = 10.5 min; *m/z* calculated for $[C_{53}H_{83}ClN_{12}O_{16}+H]^+$: 1179.58113, found: 1179.57942.

$(BG)_2$-linker-OH ($(snap)_2$)



$(BG)_2$-linker-OH

Figure S2. Structure of $(BG)_2$-linker-OH ($(snap)_2$).

All reaction steps were performed at room temperature on a peptide shaker at 1000 rpm. Unless indicated otherwise, washing refers to washing 3 × each with 5 ml NMP/DCM (1:1), then DCM, then NMP.

In a 10 ml syringe with a polyethylene frit, 75 mg (40.6 µmol, 1.00 eq) H-Gly-2-chlorotrityl resin were swelled in NMP for 1 h. 120 mg (203 µmol, 5.00 eq) FmocLys(Fmoc)-OH, 69 mg (183 µmol, 4.50 eq) HBTU, 27 mg (203 µmol, 5.00 eq) and 240 µl (184 mg, 1.42 mmol, 35.0 eq) DIPEA in 1.5 ml NMP were added to the resin and shaken for 1 h. After washing, 3 × 6 ml 20 % piperidine in NMP were applied for 10 min each for Fmoc deprotection and the resin was washed again. Then, 125 mg (325 µmol, 8.00 eq) Fmoc-AEEA-COOH was coupled to the resin with 111 mg (292 µmol, 7.20 eq) HBTU, 44 mg (325 µmol, 8.00 eq) HOBt and 387 µl (294 mg, 2.27 mmol, 56.0 eq) DIPEA in 1.5 ml NMP for 1 h. After washing, 3 × 6 ml 20 % piperidine in NMP were applied for 10 min each for Fmoc deprotection and the resin was washed again. 93 mg (812 µmol, 20.0 eq) glutaric anhydride was coupled with 138 µl (105 mg, 812 µmol, 20.0 eq) DIPEA in 1 ml NMP for 30 min. The resin was washed again with a subsequent additional washing step with 1 % NaOH in dioxane/H_2O (1:1) for 1 min and thereafter washing 4 × each with DCM/NMP (1:1), DCM and NMP. For activation of the glutaric acid, 2 × 146 µl (238 mg, 850 µmol, 21.0 eq) Pfp-OTFA in 3.75 ml pyridine/DCM (1:1) were applied for 10 min and 20 min respectively. After washing 4 × with NMP only, 60 mg (224 µmol, 5.50 eq) BG-NH_2 in 2 ml DMSO/pyridine (20:1) were coupled to the resin overnight.

Finally, the resin was washed 4 × each with NMP/DCM (1:1), DCM, NMP and Et_2O, swelled 2 × in NMP for 5 min and washed 4 × with DCM. Cleavage from the resin was executed under continuous flow with DCM/HFIP/TFA (90:10:0.5), which was removed under reduced pressure. The resulting crude product was dissolved in 7 % buffer B supplemented with 10 % DMSO, filtered and purified via preparative HPLC (5 % B to 50 % B in 40 min), which yielded 25 mg (20.0 µmol, 36 %) $(BG)_2$-linker-OH as a colorless powder after lyophilisation.

UV spectrum (PBS): λ_{max} = 282 nm, $\varepsilon_{260\ nm}$ = 5.00 $mM^{-1}cm^{-1}$; t_R = 8.5 min; *m/z* calculated for $[C_{56}H_{75}N_{17}O_{15}+H]^+$: 1226.57013, found: 1226.57121.

(Chloroalkane)$_2$-linker-OH ((halo)$_2$)

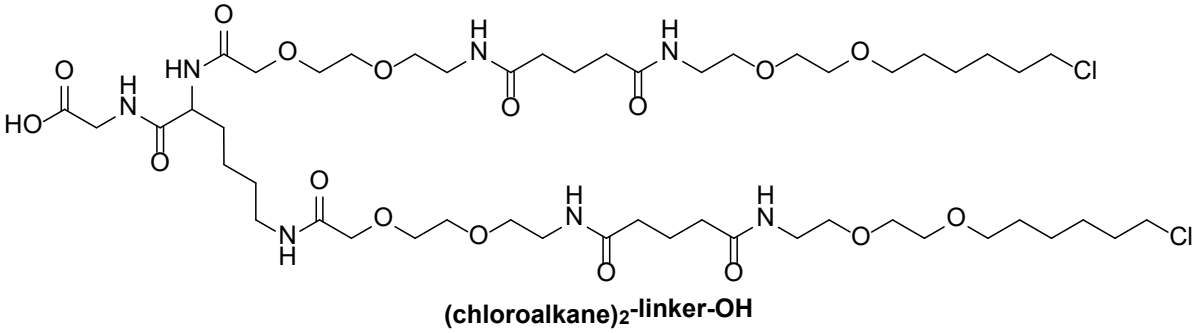


Figure S3. Structure of (chloroalkane)$_2$-linker-OH ((halo)$_2$).

(Chloroalkane)$_2$-linker-OH was obtained via the solid phase peptide synthesis protocol described for (BG)$_2$-linker-OH by coupling 50 mg (224 µmol, 5.50 eq) chloroalkane-NH_2 in 2 ml NMP instead of BG-NH_2. The resulting crude product was dissolved in 33 % buffer B, filtered and purified via preparative HPLC (35 % B to 70 % B in 40 min), which yielded 29 mg (25.7 µmol, 63 %) (chloroalkane)$_2$-linker-OH as a colorless oil after lyophilization.

t_R = 14.8 min; *m/z* calculated for $[C_{50}H_{91}Cl_2N_7O_{17}+Na]^+$: 1154.57407, found: 1154.57269.

TMR-chloroalkane

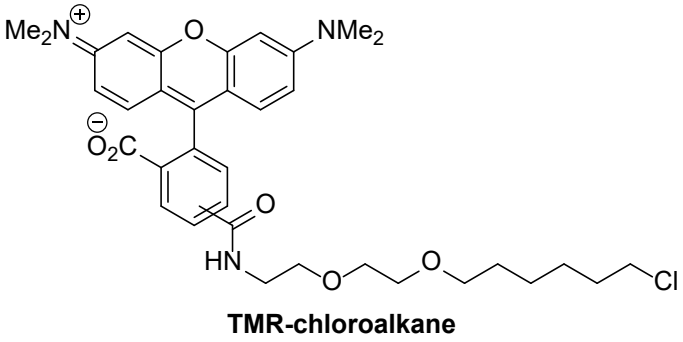
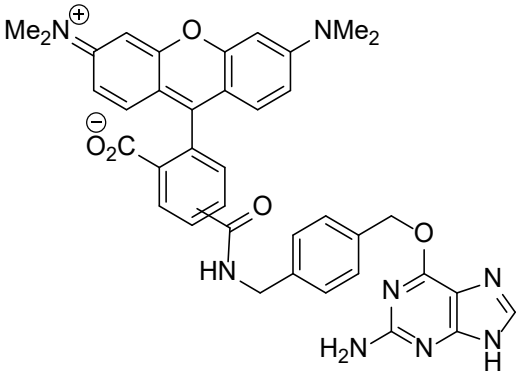


Figure S4. Structure of TMR-chloroalkane.

N-(2-(2-(6-Chloro-hexyloxy)-ethoxy)-ethyl)-tetramethylrhodamine-5(6)-amide (TMR-chloroalkane) was synthesized starting from chloroalkane-NH_2 and 5(6)-carboxytetramethylrhodamine *N*-succinimidyl ester according to literature.[7]

TMR-BG



TMR-BG

Figure S5. Structure of TMR-BG.

TMR-BG was obtained via the protocol described for TMR-chloroalkane[7] by using BG-NH_2 instead of chloroalkane-NH_2 and replacing the solvent with DMSO.

UV spectrum (H_2O/MeCN): λ_{max} = 552 nm, $\varepsilon_{543\,nm}$ = 95.826 $mM^{-1}cm^{-1}$; t_R = 11.2 min; LCMS *m/z* found for $[C_{38}H_{34}N_8O_5+2H]^{2+}$: 342.35.

Generation of guideRNAs

As guideRNAs, 22 nt long RNAs with a 5'-C6-aminolinker (NH_2-guideRNAs) that were chemically stabilized in an antagomir-like fashion as described before[8] were applied. All guideRNAs were purchased either from BIOSPRING purified via ion exchange HPLC or from EUROGENTEC purified either via reverse phase HPLC or desalted. guideRNAs that were only desalted were further purified by precipitation with 0.1 volumes of 3 M NaCl and 3.0 volumes of EtOH prior to reaction with the respective self-labeling moiety.

Sequences and extinction coefficients at 260 nm of all guideRNAs used are displayed in Table S1.

Table S1. Sequences and $\varepsilon_{260\,nm}$ of used guideRNAs. Italics = 2'OMe, s = phosphorothioate linkage, lowercase = DNA base.

guideRNA	Target	Sequence	$\varepsilon_{260\,nm}$ / $mM^{-1}cm^{-1}$
NH_2-UAG	eGFP W58X	*UsAsUGUGUCGG* CCA *CGGAAsCsAsGsG*	226.00
snap-UAG	eGFP W58X	*UsAsUGUGUCGG* CCA *CGGAAsCsAsGsG*	228.50
clip-UAG	eGFP W58X	*UsAsUGUGUCGG* CCA *CGGAAsCsAsGsG*	230.24
halo-UAG	eGFP W58X	*UsAsUGUGUCGG* CCA *CGGAAsCsAsGsG*	226.00
NH_2-UAC	GAPDH T211A	*CsCsGAGCGCCA* GCA *GAGGCsAsGsGsG*	222.00
snap-UAC	GAPDH T211A	*CsCsGAGCGCCA* GCA *GAGGCsAsGsGsG*	224.50
halo-UAC	GAPDH T211A	*CsCsGAGCGCCA* GCA *GAGGCsAsGsGsG*	222.00
NH_2-UAU	GAPDH I30V	*CsUsAGGCAACA* ACA *UCCACsUsUsUsA*	224.00
snap-UAU	GAPDH I30V	*CsUsAGGCAACA* ACA *UCCACsUsUsUsA*	226.50
halo-UAU	GAPDH I30V	*CsUsAGGCAACA* ACA *UCCACsUsUsUsA*	224.00
NH_2-CAG	ACTB S323G	*GsAsACAUUGUG* CCG *GGUGCsCsAsGsG*	214.70
snap-CAG	ACTB S323G	*GsAsACAUUGUG* CCG *GGUGCsCsAsGsG*	217.20
halo-CAG	ACTB S323G	*GsAsACAUUGUG* CCG *GGUGCsCsAsGsG*	214.70
halo-snap-CAG	ACTB S323G	*GsAsACAUUGUG* CCG *GGUGCsCsAsGsG*	217.20
$(snap)_2$-CAG	ACTB S323G	*GsAsACAUUGUG* CCG *GGUGCsCsAsGsG*	219.70

(halo)$_2$-CAG	ACTB S323G	*GsAsACAUUGUG* CCG *GGUGCsCsAsGsG*	214.70
NH$_2$-CAG_2	GAPDH T177T	*UsAsCGCAUGGA* CCG *UGGUCsAsUsGsA*	226.00
snap-CAG_2	GAPDH T177T	*UsAsCGCAUGGA* CCG *UGGUCsAsUsGsA*	228.50
halo-CAG_2	GAPDH T177T	*UsAsCGCAUGGA* CCG *UGGUCsAsUsGsA*	226.00
NH$_2$-CAU	GAPDH I38V	*CsAsAGAGGUCA* ACG *AAGGGsGsUsCsA*	238.00
snap-CAU	GAPDH I38V	*CsAsAGAGGUCA* ACG *AAGGGsGsUsCsA*	240.50
halo-CAU	GAPDH I38V	*CsAsAGAGGUCA* ACG *AAGGGsGsUsCsA*	238.00
halo-snap-CAU	GAPDH I38V	*CsAsAGAGGUCA* ACG *AAGGGsGsUsCsA*	240.50
(snap)$_2$-CAU	GAPDH I38V	*CsAsAGAGGUCA* ACG *AAGGGsGsUsCsA*	243.00
(halo)$_2$-CAU	GAPDH I38V	*CsAsAGAGGUCA* ACG *AAGGGsGsUsCsA*	238.00
NH$_2$-eGFP	eGFP W58X + T63M	*GsUsGUAGUGUCGG* CCA *CGGAAsCsAsGsG*	249.00
snap-eGFP	eGFP W58X + T63M	*GsUsGUAGUGUCGG* CCA *CGGAAsCsAsGsG*	251.50
halo-eGFP	eGFP W58X + T63M	*GsUsGUAGUGUCGG* CCA *CGGAAsCsAsGsG*	249.00
halo-snap-GFP	eGFP W58X + T63M	*GsUsGUAGUGUCGG* CCA *CGGAAsCsAsGsG*	251.00
mod-NH$_2$-eGFP	eGFP T63M	*UUUUAGUGUCGGCCACGGAACAGG*	238.6
mod-snap-eGFP	eGFP T63M	*UUUUAGUGUCGGCCACGGAACAGG*	241.1
NH$_2$-UAU	ACTB I5V	*CsUsCCGCGGCG* ACA *UCAUCsAsUsCsC*	199.60
(halo)$_2$-UAU	ACTB I5V	*CsUsCCGCGGCG* ACA *UCAUCsAsUsCsC*	199.60
NH$_2$-ACC	GAPDH T52I	*CCUAUCAUAUUGGAACAUGUAAAC*	247.20
(snap)$_2$-ACC	GAPDH T52I	*CCUAUCAUAUUGGAACAUGUAAAC*	252.20
anti-(snap)$_2$-ACC		atggtttacatgttccaatatgatagg	
snap-UAG_2	ACTB 3'-UTR	*UsCsGAGCAAUG* CCA *UCACCsUsCsCsC*	209.50

Constructs for stable cell lines

The sequences of the respective constructs are attached as Appendix. Furthermore, full plasmid maps with assigned features and restriction sites are additionally supplied as SnapGene files.

Single cell lines

SA1Q, CA1Q and HA1Q were cloned in a pcDNA 5 vector under control of the CMV promoter followed by two copies of the *tet* operator (TetO$_2$) via restriction/ligation (BamHI/NotI, NEW ENGLAND BIOLABS). C-terminally, a Myc- and a His-tag, followed by the targeted UAG codon in the 3'-UTR, were attached.

HA1Q / SA2Q duo cell lines 1 – 5

Constructs for HA1Q / SA2Q duo cell lines **1** – **5** were cloned in a pcDNA 5 vector via restriction/ligation (BamHI/ApaI/NotI/ClaI for **1**, **2**; BamHI/PacI for **3**; ClaI/NotI/BamHI/PacI for **4**, **5**).

HA1Q / APO1S duo cell lines 6 – 9

Constructs for HA1Q / APO1S duo cell lines **6** – **9** were cloned in a pcDNA 5 vector via restriction/ligation (BamHI/ApaI/NotI/ClaI for **6**, **7**; ClaI/NotI/BamHI/PacI for **8**, **9**).

Immunostaining of single cell lines

Unless stated otherwise, incubation steps were performed at room temperature.

For immunostaining, coverslips in 4 wells of a 24 well plate were coated with 500 µl poly-D-lysine hydrobromide (0.1 mg/ml in Millipore water) for 30 min. After washing with 500 µl Millipore water and 500 µl PBS, the well plate was irradiated with a UV lamp for 30 min and subsequently allowed to dry for further 30 min. $1.2 \cdot 10^5$ SA1Q or HA1Q 293 Flp-In T-REx cells were seeded in 500 µl DMEM/FBS/B/H for – Dox samples or 500 µl DMEM/FBS/B/H/10 ng/ml doxycycline (DMEM/FBS/B/H/10 D) for + Dox samples respectively.

After 24 h, medium was removed and coverslips were washed wit 500 µl PBS. Cells were incubated with 500 µl 3.7 % formaldehyde in PBS for 20 min and washed 3 × with 500 µl PBS. For permeabilization, 500 µl 1 % Triton X-100 in PBS were added, incubated for 5 min and washed 3 × with 500 µl PBS. Then, cells were incubated with 500 µl 10 % FBS in PBS for 1.5 h and subsequently with 200 µl mouse α-Myc (1:1.000 in 10 % FBS in PBS, SIGMA ALDRICH M4439) for 2 h. Washing with 3 × 500 µl PBS was followed by incubation with 250 µl goat α-mouse Alexa Fluor 488 (1:1.000 in 10 % FBS in PBS, THERMO FISHER SCIENTIFIC A11001) for 1 h. After washing with 2 × 500 µl PBS, nuclei were stained with 200 µl NucBlue™ Live ReadyProbes™ Reagent Hoechst33342 (1:100 in PBS, THERMO FISHER SCIENTIFIC R37605) for 30 min and washed again with 2 × 500 µl PBS. Coverslips were then mounted to object slides with Fluorescence Mounting Medium by DAKO and dried overnight at 4 °C. Microscopy was performed with a ZEISS AXIO Observer.Z1 with a Colibri.2 light source under 63× magnification. For excitation and emission wavelengths, see *Table S2*.

Table S2. Excitation and emission wavelengths λ.

Channel	**green**		**blue**		**red**	
	λ	band pass filter	λ	band pass filter	λ	band pass filter
Excitation	488 nm	460 – 488 nm	353 nm	350 – 390 nm	587 nm	567 – 602 nm
Emission	509 nm	500 – 557 nm	465 nm	402 – 448 nm	610 nm	615 – 4095 nm

FITC-BG & TMR-chloroalkane staining of duo cell lines

Unless stated otherwise, incubation steps were performed at room temperature.

For FITC-BG and TMR-chloroalkane staining, coverslips in 10 wells of a 24 well plate were coated with 500 µl poly-D-lysine hydrobromide (0.1 mg/ml in Millipore water) for 30 min. After washing with 500 µl Millipore water and 500 µl PBS, the well plate was irradiated with a UV lamp for 30 min and subsequently allowed to dry for further 30 min. $5 \cdot 10^4$ 293 Flp-In T-REx cells from cell lines **1** – **5** were seeded in 500 µl DMEM/FBS/B/H for – Dox samples or 500 µl DMEM/FBS/B/H/10 D for + Dox samples respectively.

After 24 h, 303 µl of the medium were removed and replaced with 1 µl DMEM/FBS/B/H containing 0.4 mM FITC-BG and 1.0 mM TMR-chloroalkane and 2 µl NucBlue™ Live ReadyProbes™ Reagent Hoechst33342 (THERMO FISHER SCIENTIFIC R37605). After incubation for 30 min, 21.6 µl 37 % aqueous formaldehyde were added, cells were incubated for 3 min and subsequently washed 3 × with 200 µl PBS. For permeabilization, cells were incubated with 200 µl 0.1% Triton X-100 in PBS for 15 min and washed 3 × with 200 µl PBS. Coverslips were then mounted to object slides with Fluorescence Mounting Medium by DAKO and dried overnight at 4 °C. Microscopy was performed with a ZEISS AXIO Observer.Z1 with a Colibri.2 light source under 63× magnification. For excitation and emission wavelengths, see *Table S2*.

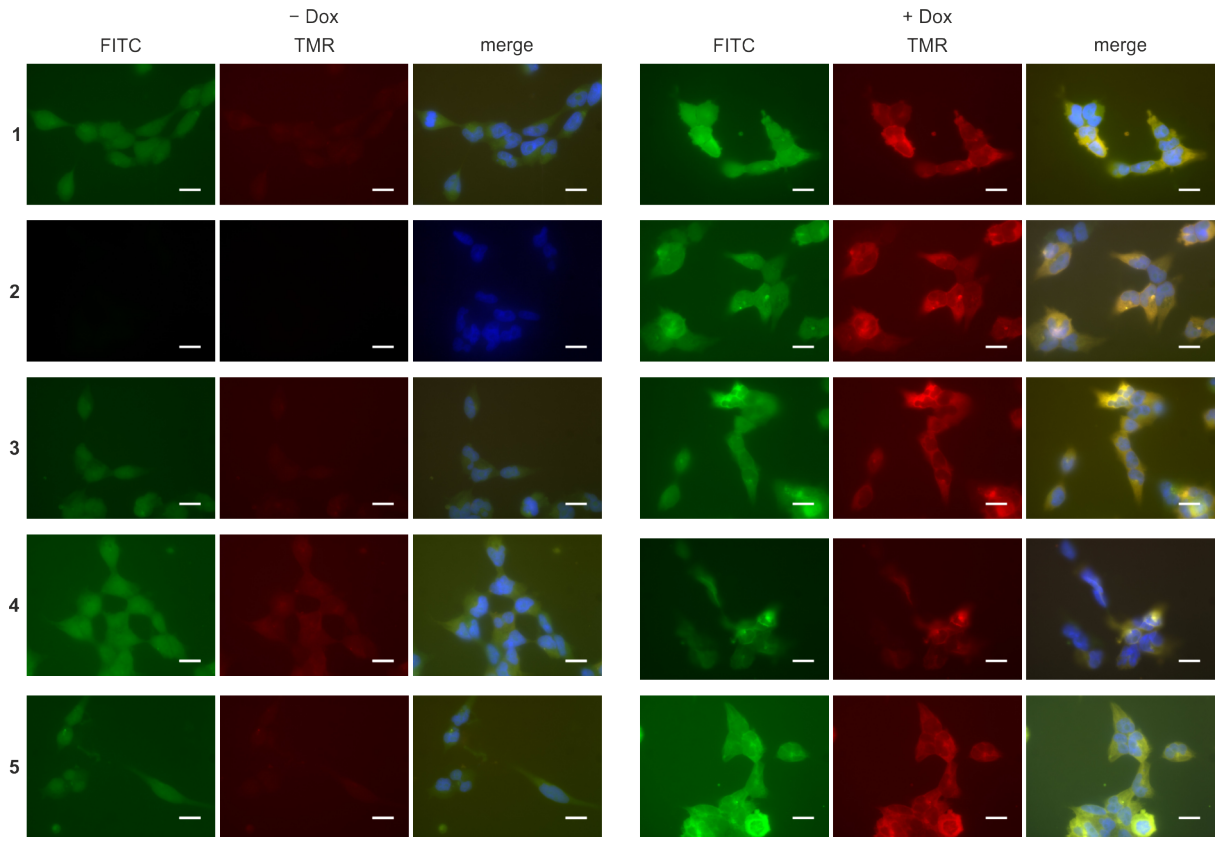


*Figure S6. FITC-BG (green channel) and TMR-chloroalkane (red channel) staining of cell lines **1** – **5** without (left panel) and with (right panel, same as Fig. 2b) doxycycline induction. Cell nuclei are stained with Hoechst33342 (blue channel).*

SDS-PAGE & Western Blotting

Expression in single cell lines

$1 \cdot 10^5$ SA1Q or HA1Q 293 Flp-In T-REx cells respectively were seeded in 500 µl medium in 2 wells per condition of a 24 well plate. For the uninduced samples (–) and the samples with 24 h of doxycycline induction (+), DMEM/FBS/B/H was used as medium; for the samples with 48 h of doxycycline induction (++) DMEM/FBS/B/H/10 D was used. After 24 h, the medium of the + samples was replaced by DMEM/FBS/B/H/10 D. After another 24 h, medium was removed and cells were washed with 500 µl PBS. After treatment with 60 µl trypsin-EDTA solution from SIGMA ALDRICH, 440 µl DMEM/FBS were added and cells were transferred to reaction tubes. Centrifugation for 5 min at 1.600 rpm was followed by washing with 500 µl PBS and resuspension of the cell pellets in 100 µl urea lysis buffer (8 M urea, 100 mM NaH_2PO_4, 10 mM Tris, pH 8.0). Cells were lysed via shear force by drawing the solution up and out a 19 gauge syringe 6×. After centrifugation for 15 min at 16.000 rpm and 4 °C, supernatants were transferred to fresh reaction tubes and total protein concentration of the samples was determined via Bradford assay from SIGMA ALDRICH.

7 µg protein lysate in 13.33 µl urea lysis buffer were heated with 4 µl 6× Laemmli buffer (0.4 M SDS, 60 mM Tris pH 6.8, 6.5 M glycerol, 0.6 M dithiothreitol, 0.9 mM bromophenol blue) for 5 min at 95 °C and 700 rpm and subsequently loaded to an SDS-PAGE, which was run at 80 V for 1.5 h followed by 120 V for 45 min. Transfer onto a PVDF membrane (BIO-RAD LABORATORIES) was performed at 30 V and 4 °C for 18 h. After blocking in 5 % dry milk in TBST containing 50 µg/ml avidin for 1 h, the blot was incubated with mouse α-Myc (1:5.000, SIGMA ALDRICH M4439) and mouse α-ACTB (1:40.000, SIGMA-

Aldrich A5441) in 5 % dry milk-TBST for 2 h at room temperature as primary antibodies. As secondary antibody, goat α-mouse HRP (1:10.000, JACKSON IMMUNORESEARCH 115-035-003) with added Precision Protein StrepTactin HRP conjugate (for visualisation of the Precision Plus Western C Standard, 1:25.000, BIO-RAD) in 5 % dry milk-TBST was applied for 2 h at room temperature. Chemiluminescence was measured with a FUSION FX by VILBER.

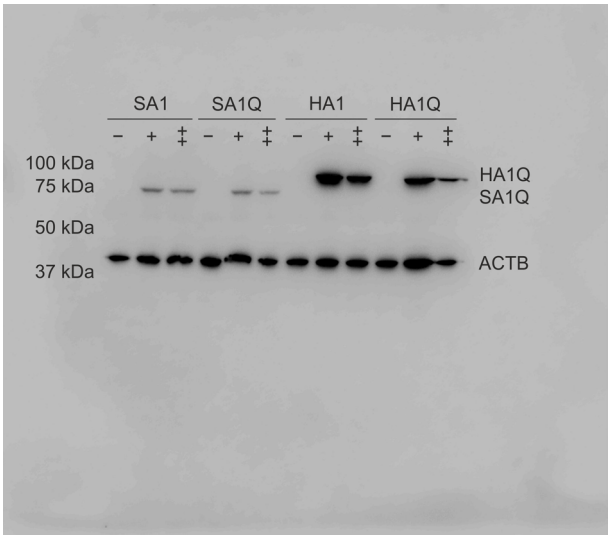


Figure S7. Full Western Blot (from Fig. 1c) of SA1, SA1Q, HA1 and HA1Q 293 Flp-In T-REx cells without (–), with 24 h (+) or 48 h (++) doxycycline induction. The different ADAR fusions were detected via a C-terminally attached Myc-tag, ACTB served as loading control.

GuideRNA-protein conjugation assay

For detection of the respective editing enzyme-guideRNA conjugate, SA1Q and HA1Q 293 Flp-In T-REx cells were transfected with varying amounts of snap- or halo-guideRNA, respectively, and characterized via Western Blot. For comparison, an additional serial dilution of the editing enzyme without guideRNA and an uninduced control sample (–) were loaded side by side. The experiment was conducted completely analogous to editing experiments of endogenous targets, differing only in the 5× greater scale to ensure sufficient protein amounts for Western Blot detection.

$2 \cdot 10^6$ SA1Q or HA1Q 293 Flp-In T-REx cells were seeded in a 6 well plate in 2.5 ml DMEM/FBS/B/H for – Dox samples or 2.5 ml DMEM/FBS/B/H/10 D for + Dox samples respectively. After 24 h, $4 \cdot 10^5$ cells were reverse transfected in a 24 well plate with 5.0, 25, 50 or 125 pmol (corresponding to 1.0, 5.0, 10 or 25 pmol per $8 \cdot 10^4$ cells on the editing experiment's scale) snap- or halo-ACC respectively with 2.5 µl Lipofectamine 2000. Doxycycline concentration for + Dox samples was kept at 10 ng/ml and after further 24 h medium was removed and cells were washed with 500 µl PBS. Cells were lysed with 70 µl 1× Laemmli (67 mM SDS, 10 mM Tris pH 6.8, 1.1 M glycerol, 0.10 M dithiothreitol, 0.15 mM bromophenol blue) in RIPA Lysis and Extraction Buffer (1 % NP-40, 150 mM NaCl, 25 mM Tris·HCl pH 7.6, 1 % sodium deoxycholate, 0.1 % SDS, THERMO FISHER SCIENTIFIC; supplemented with 1 tablet cOmplete™ Mini EDTA-free Protease Inhibitor Cocktail by ROCHE per 10 ml) and cell lysates were immediately frozen in liquid nitrogen. For SDS-PAGE, protein lysates were heated for 15 min at 95 °C and 1500 rpm and 20 µl of the respective lysate or the indicated dilution in 1× Laemmli in RIPA Lysis and Extraction buffer were loaded and run at 90 V for 5 min followed by 200 V for 110 min. Transfer onto a PVDF membrane (BIO-RAD LABORATORIES) was performed at 35 V and 4 °C for 16 h. After blocking in 5 % dry milk in TBST for 1 h, the blot was incubated with first rabbit α-ADAR1 (1:1.000, BETHYL LABORATORIES A303-884) in 5 % dry milk-TBST for 2 h at room temperature and subsequently with rabbit α-GAPDH (1:1.000, CELL SIGNALING

#5174) in 5 % dry milk-TBST overnight at 4 °C as primary antibodies. As secondary antibody, goat α-rabbit HRP (1:10.000, JACKSON IMMUNORESEARCH 111-035-003) in 5 % dry milk-TBST was applied for 2 h at room temperature. Chemiluminescence was measured with an Odyssey Fc Imaging System (LI-COR).

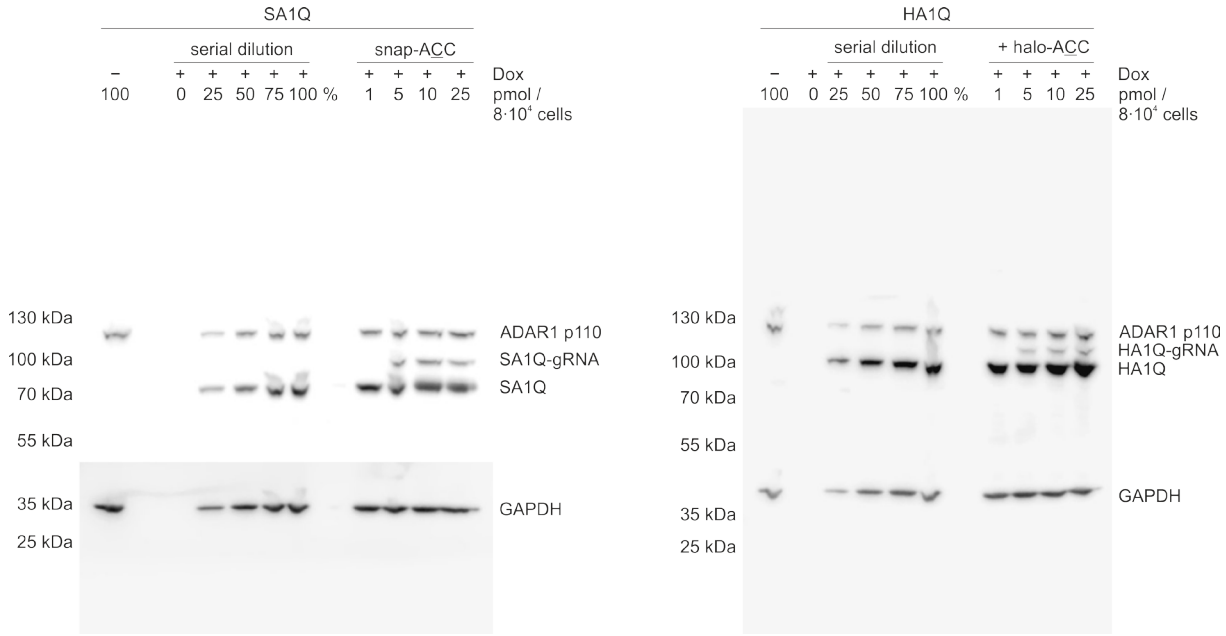


Figure S8. Full Western Blots (from Fig. 1e) of SA1Q and HA1Q 293 Flp-In T-REx cells after transfection of 1.0, 5.0, 10 or 25 pmol snap- or halo-ACC per 8·10⁴ cells, respectively. On the left, a sample without doxycycline induction (–) and a serial dilution of 0, 25, 50, 75 or 100 % lysate from cells induced with doxycycline (+), but without guideRNA are shown. For detection, an α-ADAR1 antibody was used, staining the different ADAR proteins (SA1Q, HA1Q), its guideRNA conjugates (SA1Q-gRNA, HA1Q-gRNA), but also endogenous ADAR1 p110; GAPDH served as loading control. The SA1Q blot was cut above GAPDH before detection.

Expression in duo cell lines

$2 \cdot 10^5$ 293 Flp-In T-REx cells from the respective duo cell line were seeded in 500 µl medium in 2 wells of a 24 well plate per condition. For the uninduced samples (– Dox) DMEM/FBS/B/H was used as medium, for the samples with doxycycline induction (+ Dox) DMEM/FBS/B/H/10 D was used. After 24 h, medium was removed and cells were first washed with 500 µl PBS and then detached and suspended in 500 µl fresh PBS per well. Centrifugation for 5 min at 1.600 rpm was followed by removal of PBS and resuspension of the cell pellets in 30 µl NP40 lysis buffer (1 % NP-40, 150 mM NaCl, 50 mM Tris pH 8.0; 1 tablet cOmplete™ Mini EDTA-free Protease Inhibitor Cocktail by ROCHE per 10 ml). After centrifugation for 15 min at 16.000 rpm and 4 °C, supernatants were transferred to fresh reaction tubes and total protein concentration of the samples was determined via Pierce BCA Protein Assay Kit by THERMO FISHER SCIENTIFIC.

For co-staining with TMR-BG and TMR-chloroalkane, 10 µg protein lysate was incubated with 5 µM TMR-BG and TMR-chloroalkane each in 13.33 µl NP40 lysis buffer for 30 min at 37 °C and 600 rpm. 4 µl 6× Laemmli buffer (0.4 M SDS, 60 mM Tris pH 6.8, 6.5 M glycerol, 0.6 M dithiothreitol, 0.9 mM bromophenol blue) were added, samples were heated for 5 min at 95 °C and 700 rpm and subsequently loaded to an SDS-PAGE. TMR staining on the completed SDS-PAGE was visualized on a FLA 5100 by FUJIFILM with excitation at 532 nm and emission at 557 nm (Cy3 filter set) and a resolution of 10 µm.

Transfer onto a PVDF membrane (BIO-RAD LABORATORIES) was performed at 28 V and 4 °C for 18.5 h. After blocking in 5 % dry milk in TBST containing 50 µg/ml avidin for 1 h, the blot was incubated with first mouse α-ACTB (1:40.000, SIGMA-Aldrich A5441) in 5 % dry milk-TBST for 2 h at room temperature and subsequently with rabbit α-SNAP-tag (1:1.000, NEW ENGLAND BIOLABS P9310S) and rabbit α-HaloTag (1:1.000, PROMEGA G9281) in 5 % dry milk-TBST overnight at 4 °C as primary antibodies. As secondary antibodies, first goat α-mouse HRP (1:5.000, JACKSON IMMUNORESEARCH 115-035-003) with added Precision Protein StrepTactin HRP conjugate (for visualisation of the Precision Plus Western C Standard, 1:25.000, BIO-RAD) in 5 % dry milk-TBST was applied for 2 h at room temperature, followed by goat α-rabbit HRP (1:5.000, JACKSON IMMUNORESEARCH 111-035-003) for 2 h at room temperature. Chemiluminescence was measured with a FUSION FX by VILBER.

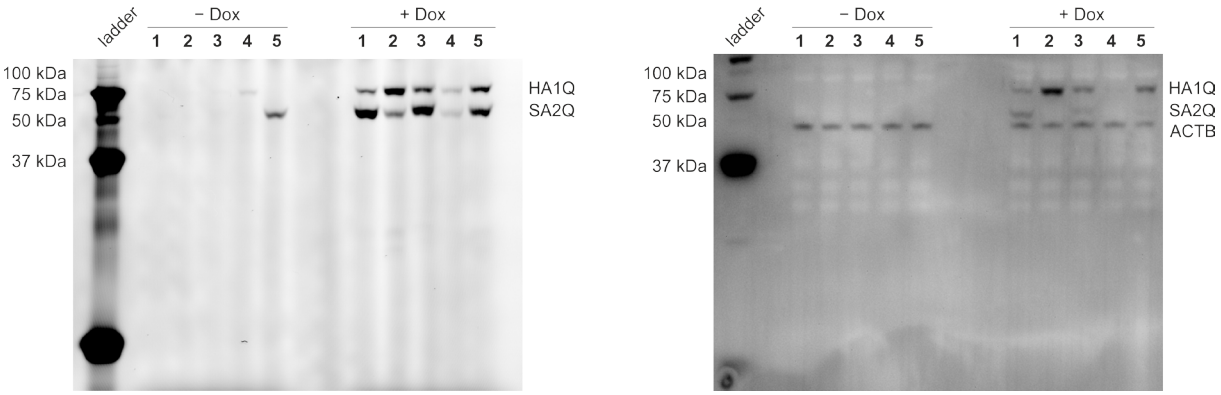


*Figure S9. TMR stained SDS-PAGE (left panel, from Fig. 2c) and Western Blot (right panel) of HA1Q / SA2Q 293 Flp-In T-REx duo cells **1** – **5** without (– Dox) and with (+ Dox) doxycycline induction. The different ADAR fusions were detected via co-staining with TMR-BG and TMR-chloroalkane or antibodies against SNAP-tag and HaloTag respectively, ACTB served as loading control.*

Editing experiments

All editing experiments depicted in bar graphs were conducted in biological triplicates and standard deviations are shown. The exact editing yields can also be found in tabular form as additional supporting file.

Editing under transient expression of editing enzymes

As a first test, editing of a premature stop codon (W58X) in eGFP with transiently expressed $SNAP_f$-, $CLIP_f$- and HALO-ADAR1 (wildtype deaminase domain) was compared.

$2 \cdot 10^5$ wildtype 293T cells were seeded in 500 µl DMEM/FBS/1 % penicillin/1 % streptomycin (DMEM/FBS/P/S) in a 24 well plate. After 24 h, medium was replaced with 450 µl DMEM/FBS and 500 ng eGFP W58X in pcDNA 3.1 plus either 100 ng $SNAP_f$-, $CLIP_f$- or HALO-ADAR1 in pcDNA 3.1 were forward transfected with 2.4 µl Lipofectamine 2000. 24 h thereafter, $6 \cdot 10^4$ cells were reverse transfected in a 96 well plate with 10 pmol of the respective guideRNA with 2.2 µl Lipofectamine 2000. After further 24 h, cells were examined under a ZEISS AXIO Observer.Z1 microscope with a Colibri.2 light source under 5× magnification for successfully edited and therefore fluorescent eGFP. For excitation and emission wavelengths, see *Table S2*. Then, cells were harvested and RNA isolation was performed with the Monarch® RNA cleanup kit from NEW ENGLAND BIOLABS, followed by DNase I digestion. eGFP RNA was reverse transcribed to cDNA, which was amplified via Taq PCR and subsequently analyzed with Sanger sequencing (either EUROFINS GENOMICS or MICROSYNTH). A-to-I

editing yields were determined by dividing the peak height for guanosine by the sum of the peak heights for both adenosine and guanosine.

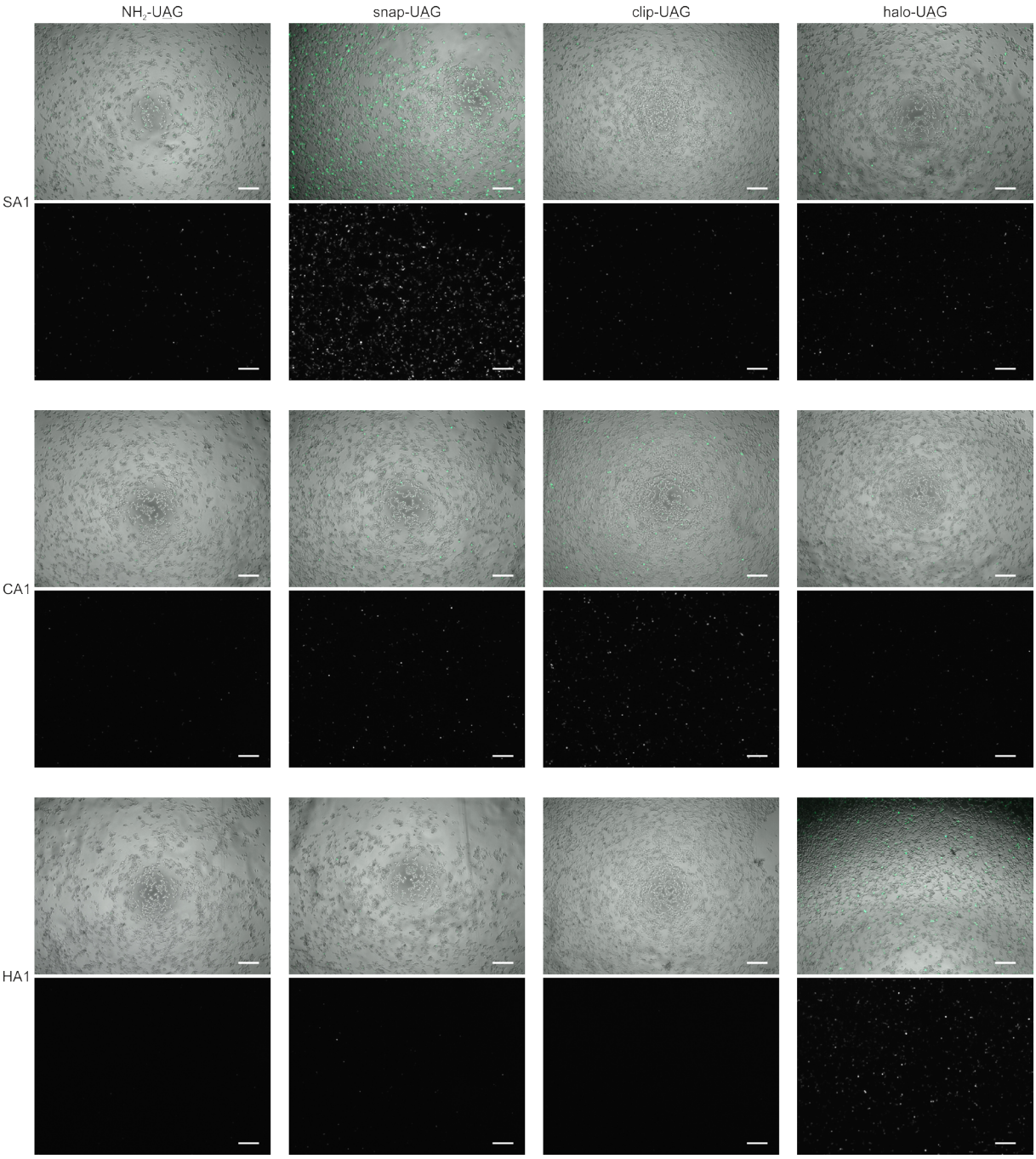


Figure S10. eGFP fluorescence (lower panels) and overlays with bright field images (upper panels) of wildtype 293T cells transiently transfected with eGFP W58X and either SA1, CA1 or HA1 after transfection of 10 pmol NH_2- (negative control), snap-, clip- or halo-UAG guideRNA. Scale bars correspond to 250 µm.

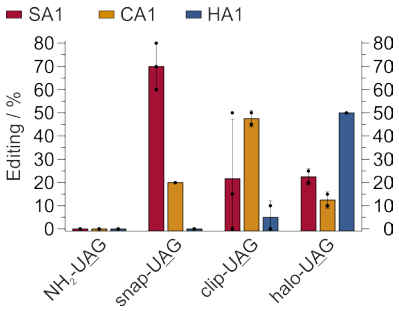


Figure S11. Editing efficiencies and orthogonality of NH_2- (negative control), snap-, clip- and halo-UAG targeting a premature W58X stop codon in eGFP in wildtype 293T cells transiently expressing either SA1, CA1 or HA1. The guideRNAs differ only in the indicated self-labeling moiety. The NH_2-guideRNA refers to a control guideRNA lacking a self-labeling moiety.

Editing of endogenous targets under genomic expression of editing enzymes

$4 \cdot 10^5$ of the respective 293 Flp-In T-REx cells were seeded in 500 µl DMEM/FBS/B/H/10 D in a 24 well plate. After 24 h, $8 \cdot 10^4$ cells were reverse transfected in a 96 well plate with the respective amount of the guideRNA to be examined with 0.5 µl Lipofectamine 2000. Doxycycline concentration was kept at 10 ng/ml and after further 24 h (or 48 h for cell lines expressing APO1S) cells were harvested. RNA isolation was performed with the Monarch® RNA cleanup kit from NEW ENGLAND BIOLABS, followed by DNase I digestion. Samples containing $(snap)_2$-ACC were treated with a DNA oligonucleotide of complementary sequence (anti-$(snap)_2$-ACC, 1 µM) at 95 °C for 3 min to trap the guideRNA. Purified RNA was then reverse transcribed to cDNA, which was amplified via Taq PCR and subsequently analyzed with Sanger sequencing (either EUROFINS GENOMICS or MICROSYNTH). A-to-I editing yields were determined by dividing the peak height for guanosine by the sum of the peak heights for both adenosine and guanosine, C-to-U editing yields by dividing the peak height for thymidine by the sum of the peak heights for both cytidine and thymidine.

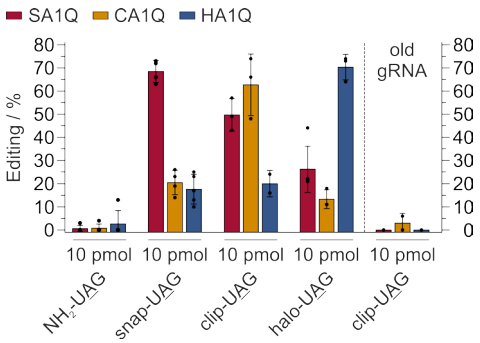


Figure S12. Editing efficiencies and orthogonality of NH_2- (negative control), snap-, clip- and halo-UAG targeting a 5´-UAG reporter codon in the 3'-UTR in three different cell lines, each expressing one ADAR1 fusion protein (SNAP-ADAR1Q (SA1Q), CLIP-ADAR1Q (CA1Q), HALO-ADAR1Q), as indicated. clip-guideRNA shows loss of activity upon long-term storage.

*Table S3. Screening of duo cell lines **1** – **5**. Maximum editing yield and selectivity after single or co-transfection of a snap- and/or a halo-guideRNA (snap-/halo-CAG_2 and snap-/halo-CAU, 5.0 pmol each) for a CAG and a CAU codon in the ORF of GAPDH.*

Cell line	**1**		**2**		**3**		**4**		**5**	
	CAG	CAU	CAG	CAU	CAG	CAU	CAG	CAU	CAG	CAU
Editing yield	50 %	10 %	35 %	35 %	45 %	15 %	7 %	5 %	45 %	30 %
Selectivity	3.3x	∞	-	∞	1.2x	6.8x	-	-	1.2x	∞

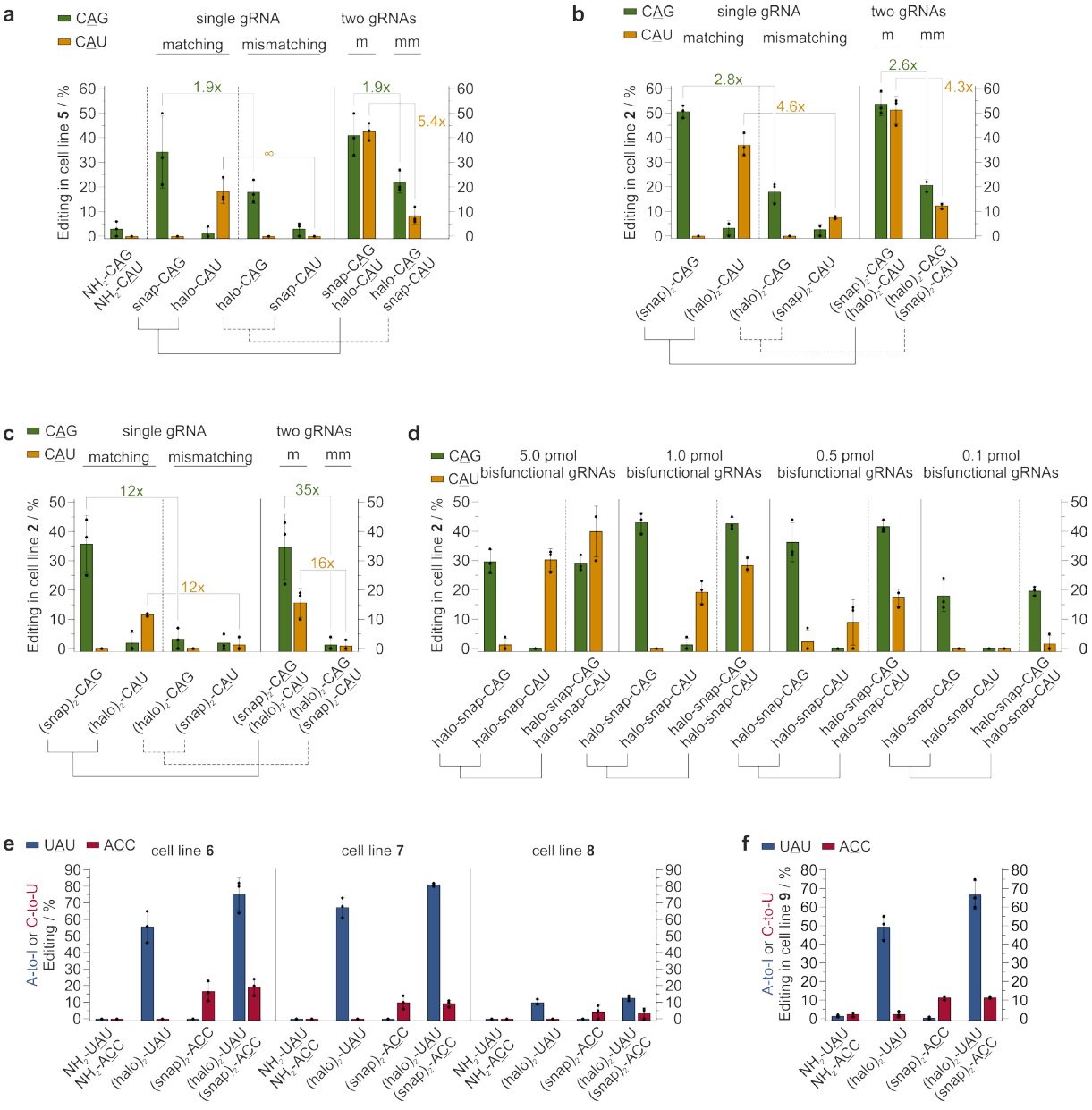


*Figure S13. **a)** Editing yield and selectivity after transfection of a single (5.0 pmol), matching or mismatching snap- or halo-guideRNA (left panel) into duo cell line **5** compared to the co-transfection of two guideRNAs (one snap- and one halo-guideRNA, each 5.0 pmol) either in matching (m) or in mismatching (mm) combination **b)** Editing yield and selectivity in duo cell line **2** after transfection of 1.0 pmol of a single $(snap)_2$- or $(halo)_2$-guideRNA or after co-transfection of a $(snap)_2$- and a $(halo)_2$-guideRNA, either in matching (m) or mismatching (mm) combination respectively. **c)** Same as b) but*

with 0.1 pmol guideRNAs. ***d)*** *Concentration dependency of editing yield in duo cell line* ***2*** *after transfection of either a single or two (same as Fig. 3h) bisfunctional halo-snap-guideRNAs.* ***e)*** *Editing yield in duo cell lines* ***6****,* ***7*** *and* ***8*** *after transfection of a single or cotransfection of two guideRNAs, one* $(halo)_2$*-guideRNA for A-to-I editing in ACTB and one* $(snap)_2$*-guideRNA for C-to-U editing in GAPDH (5.0 pmol each). As expected from the transgene expression, cell line* ***8*** *(construct analog to cell line* ***4****) shows only minor editing.* ***f)*** *Same as e) but with 2.5 pmol guideRNAs in duo cell line* ***9****.*

Editing of a transfected reporter transcript under genomic expression of editing enzymes

$2 \cdot 10^5$ of the respective 293 Flp-In T-REx cells were seeded in 500 µl DMEM/FBS/B/H/10 D in a 24 well plate. 24 h thereafter, each well was forward transfected with 300 ng pcDNA 3.1 containing the coding sequence for eGFP-W58X with 1.2 µl Lipofectamine 2000. After 24 h, $8 \cdot 10^4$ cells were reverse transfected in a 96 well plate with the respective amount of the guideRNA to be examined with 0.5 µl Lipofectamine 2000. Doxycycline concentration was kept at 10 ng/ml and after further 48 h cells were harvested. RNA isolation was performed with the Monarch® RNA cleanup kit from NEW ENGLAND BIOLABS, followed by DNase I digestion. RNA was then reverse transcribed to cDNA, which was amplified via Taq PCR and subsequently analyzed with Sanger sequencing (either EUROFINS GENOMICS or MICROSYNTH). A-to-I editing yields were determined by dividing the peak height for guanosine by the sum of the peak heights for both adenosine and guanosine, C-to-U editing yields by dividing the peak height for thymidine by the sum of the peak heights for both cytidine and thymidine.

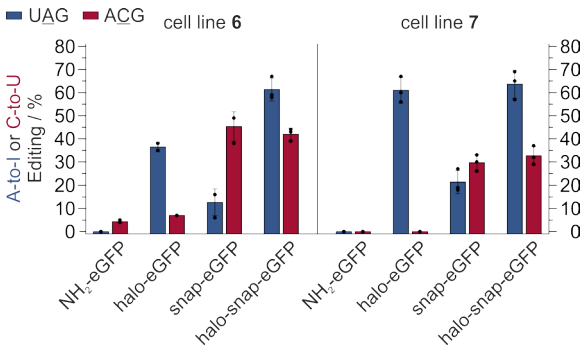


Figure S14. Editing yield in cell lines ***6*** *and* ***7*** *from concurrent A-to-I and C-to-U editing in an eGFP reporter transcript after transfection of a halo-, snap- or halo-snap-guideRNA (5.0 pmol).*

Benchmark with RESCUE

For guideRNA expression in RESCUE editing experiments, DNA oligonucleotides (Table S4) were golden-gate cloned into the guideRNA expression vector (Addgene #103852) as previously described.[9] As per requirement of the U6 promoter, a 5'-G was added in case the sequence did not start with one.

Table S4. Sequences and C or U flip positions of guideRNAs applied for editing with RESCUE. Sequences are shown in 5'-orientation. For cloning, the listed sequences and the complementary strands were preceded by a 5'-CACC or 5'-CAAC respectively.

Flip position	Target	Sequence
C flip 26	eGFP T63M	gcgt cct cactagtgtcggccacggaacagg
C flip 24	eGFP T63M	gagcgt cct cactagtgtcggccacggaaca
C flip 22	eGFP T63M	gagagcgt cct cactagtgtcggccacggaa
C flip 20	eGFP T63M	ggcagagcgt cct cactagtgtcggccacgg

C flip 26	GAPDH T52I	gatg gct ggaatcatattggaacatgtaaac
C flip 24	GAPDH T52I	gccatg gct ggaatcatattggaacatgtaa
C flip 22	GAPDH T52I	gtgccatg gct ggaatcatattggaacatgt
C flip 20	GAPDH T52I	gtttgccatg gct ggaatcatattggaacat
U flip 24	PPIB R7C	gtgttg ctt tcggagaggcgcagcatccaca
C flip 24	PPIB R7C	gtgttg cct tcggagaggcgcagcatccaca
C flip 22	PPIB R7C	gcatgttgc cct cggagaggcgcagcatcca
C flip 20	PPIB R7C	gttcatgttgc cct cggagaggcgcagcatc

Editing experiments were conducted as previously described.[9] Briefly, $2 \cdot 10^4$ 293FT cells were seeded in 150 µl DMEM/FBS in a 96 well plate. 16 h thereafter, cells were forward transfected with 150 ng RESCUEr16 expression vector (Addgene #130661), 300 ng corresponding guideRNA expression vector (sequences see Table S4) and 40 ng eGFP in pcDNA 3.1 with 0.5 µl Lipofectamine 2000. To ensure equal treatment of cells, eGFP in pcDNA 3.1 was also transfected to editing experiments targeting endogenous transcripts. After 48 h, cells were harvested and RNA isolation was performed with the Monarch® RNA cleanup kit from NEW ENGLAND BIOLABS, followed by DNase I digestion. RNA was then reverse transcribed to cDNA, which was amplified via Taq PCR and subsequently analyzed with Sanger sequencing (either EUROFINS GENOMICS or MICROSYNTH). C-to-U editing yields were determined by dividing the peak height for thymidine by the sum of the peak heights for both cytidine and thymidine, A-to-I bystander editing yields by dividing the peak height for guanosine by the sum of the peak heights for both adenosine and guanosine.

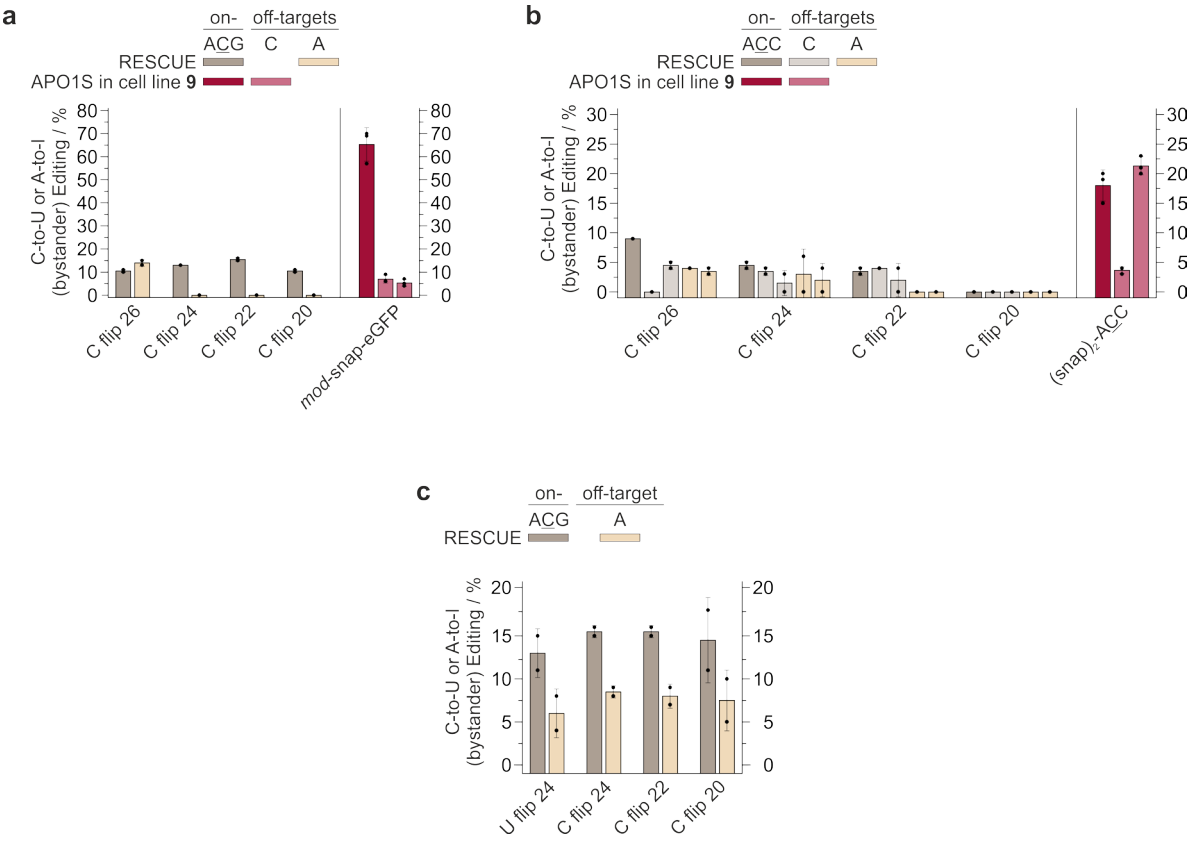


Figure S15. *Comparison of C-to-U editing yield on the same target C with RESCUEr16[9] vs. APO1S in cell line **9**. For RESCUE, four different guideRNA designs with a C or U flip at the indicated position were tested for each target, as recommended by Abudayyeh et al.[9] **a)** Editing of an eGFP reporter transcript*

with C flip 26 – 20 (300 ng) for RESCUEr16 and mod-snap-eGFP (5.0 pmol) for APO1S. ***b)*** *Editing of endogenous GAPDH with C flip 26 – 20 for RESCUEr16 (300 ng) and (snap)$_2$-ACC (5.0 pmol) for APO1S.* ***c)*** *Control for RESCUEr16 with editing in endogenous PPIB with U or C flip 24 – 20 (300 ng). The positions of the respective bystander off-targets are summarized in Table S5. Data shown as the mean ± s.d. of N = 2 – 3 independent experiments.*

Table S5. Positions of bystander off-targets in C-to-U editing with RESCUEr16 or APO1S in cell line ***9*** *from Figure S15.*

Target	Editase	Off-target	Distance to on-target / bp
eGFP T63M	RESCUE	A	– 5
	APO1S in cell line **9**	C #1	+ 30
		C #2	+ 46
GAPDH T52I	RESCUE	C #1	+ 1
		C #2	+ 21
		A #1	– 9
		A #2	– 6
	APO1S in cell line **9**	C #1	+ 21
		C #2	+ 253
PPIB R7C	RESCUE	A	– 4

Next generation sequencing

HA1Q / SA2Q duo cell line

For NGS, four samples were prepared, i.e. a duplicate of an empty transfection and a duplicate of a guideRNA transfection, both in doxycycline-induced cell line **2**. $2 \cdot 10^6$ cells from cell line **2** were seeded in 2.5 ml DMEM/FBS/B/H/10 D in a 6 well plate. After 24 h, $8 \cdot 10^4$ cells per well were reverse transfected in 5 wells of a 96 well plate per sample. For the duplicate of the empty transfection, the cells were treated with an empty reverse transfection with 0.5 µl Lipofectamine 2000 only. For the duplicate of the guideRNA transfection, cells were reverse transfected with 0.5 pmol (snap)$_2$-CAG and 0.5 pmol (halo)$_2$-CAU with 0.5 µl Lipofectamine 2000. Doxycycline concentration was kept at 10 ng/ml and after further 24 h cells were harvested. RNA isolation was performed with the RNeasy MinElute Cleanup Kit from QIAGEN, followed by DNase I digestion, which was again purified via RNeasy MinElute Cleanup Kit from QIAGEN. mRNA next generation sequencing was then performed by CEGAT. The library was prepared with the library preparation kit TruSeq Stranded mRNA by ILLUMINA starting from 100 ng RNA. Samples were then sequenced on a NovaSeq 6000 by ILLUMINA with 50 million reads and 2 × 100 bp paired end.

For comparison, data previously generated by Vogel *et al.*[10] was reanalyzed with the more sensitive pipeline applied here. Briefly, for the duplicate editing experiments of these samples, 5.0 pmol of a guideRNA targeting a 5'-UAG codon in the 3'-UTR of ACTB (snap-UAG_2, see Table S1) were transfected into 293 Flp-In T-REx cell lines either expressing SA1Q only (GSM3083480, SA1Q_rep1 & GSM3083481, SA1Q_rep2) or SA2Q only (GSM3083482, SA2Q_rep1 & GSM3083483, SA2Q_rep2). As negative control, an empty 293 Flp-In T-REx cell line expressing no artificial editing enzyme, treated with an empty transfection of Lipofectamine 2000 only (GSM3083474, ctrl_rep1 & GSM3083475, ctrl_rep2), was applied for all data sets. Data analysis is described in the Materials & Methods section.

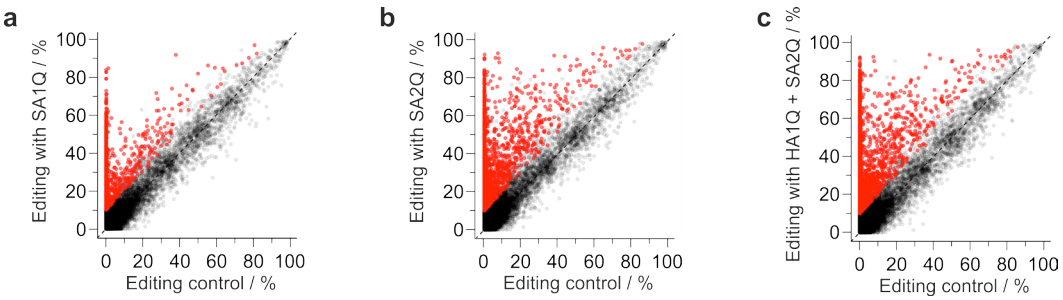


Figure S16. Scatter plots of total off-targets in editing experiments. Significantly differently edited sites are marked in red. ***a)*** *Cell line expressing SA1Q only after transfection of 5.0 pmol snap-UAG_2 versus empty 293 Flp-In T-REx.*[10] ***b)*** *Cell line expressing SA2Q only after transfection of 5.0 pmol snap-UAG_2 versus empty 293 Flp-In T-REx.*[10] ***c)*** *Cell line* ***2*** *(HA1Q + SA2Q) after transfection of 0.5 pmol $(snap)_2$-CAG and 0.5 pmol $(halo)_2$-CAU versus empty 293 Flp-In T-REx.*

Table S6. Number of significantly differently edited sites with editing difference ≥ 25 % found in editing experiments in mono cell lines SA1Q, SA2Q,[10] *and in duo cell line* ***2*** *(HA1Q + SA2Q) in comparison to a negative control cell line (293 Flp-In T-REx) not expressing any editing enzyme (Total off-targets with Δ ≥ 25 %). The last column shows the guideRNA-dependent fraction of the total off-targets with editing difference ≥ 25 % for duo cell line* ***2****.*

	Total off-targets with Δ ≥ 25 %			**gRNA-depend.**
	SA1Q	**SA2Q**	**HA1Q + SA2Q**	**HA1Q + SA2Q**
Total number	706	1423	2444	37
incl. Alu sites	75	411	418	14
5'UTR	25	46	72	1
Missense mutation	134	260	556	5
Nonstop mutation	9	10	20	0
Start codon SNP	0	0	2	0
Silent	102	143	314	0
3'UTR	332	662	1081	18
Noncoding	104	302	399	13

*Table S7. guideRNA-dependent off-target sites found by NGS in cell line **2**. Listed are significantly differently edited sites with an editing difference ≥ 25 % between duo cell line **2** with guideRNAs versus without guideRNAs. On-target sites targeted by (snap)$_2$-CAG and (halo)$_2$-CAU are shown on the top.*

Entry no.	Site	Localization	Editing / %		
			without gRNAs	with gRNAs	difference
On-target	ACTB	Missense Mutation	0	52	52
	GAPDH	Missense Mutation	0	41	41
1	UBA52	3'UTR	0	56	56
2	ACTA2	Missense Mutation	0	47	47
3	FAM50A	3'UTR	3	44	40
4	EGFL7	3'UTR	0	38	38
5	HNRNPA1L2	Noncoding	20	56	36
6	KLHDC3	3'UTR	24	58	34
7	LARP6	Noncoding	0	32	32
8	FAM129A	3'UTR	18	50	32
9	MAN2B2	Missense Mutation	5	37	32
10	UBIAD1	Noncoding	18	50	32
11	SYNGR1	Noncoding	0	31	31
12	KCNJ14	Noncoding	46	76	30
13	Unknown	Noncoding	20	50	30
14	COL4A1	3'UTR	0	30	30
15	RP13-36G14.4	Noncoding	23	53	30
16	TTC33	Noncoding	3	32	29
17	PAQR5	3'UTR	37	66	28
18	CDC42BPB	Missense Mutation	18	46	28
19	GOLGA8A	Noncoding	19	47	28
20	XRCC2	5'UTR	51	79	28
21	PVR	Noncoding	40	68	28
22	UGGT1	3'UTR	13	41	28
23	MTRF1L	Noncoding	28	55	27
24	HADHA	3'UTR	0	27	27
25	SCARB1	3'UTR	4	30	26
26	TMEM17	3'UTR	18	45	26
27	RP13-36G14.4	Noncoding	41	67	26
28	SYPL2	3'UTR	33	59	26
29	EPHB2	3'UTR	11	37	26
30	RCOR1	3'UTR	32	58	26
31	ZNF101	3'UTR	3	29	26
32	RPL28	3'UTR	28	54	26
33	WAC-AS1	Noncoding	38	64	26
34	PPDPF	3'UTR	0	26	26
35	MYL6B	Missense Mutation	0	26	26
36	ARHGAP44	Missense Mutation	10	35	25
37	SLC25A48	3'UTR	0	25	25

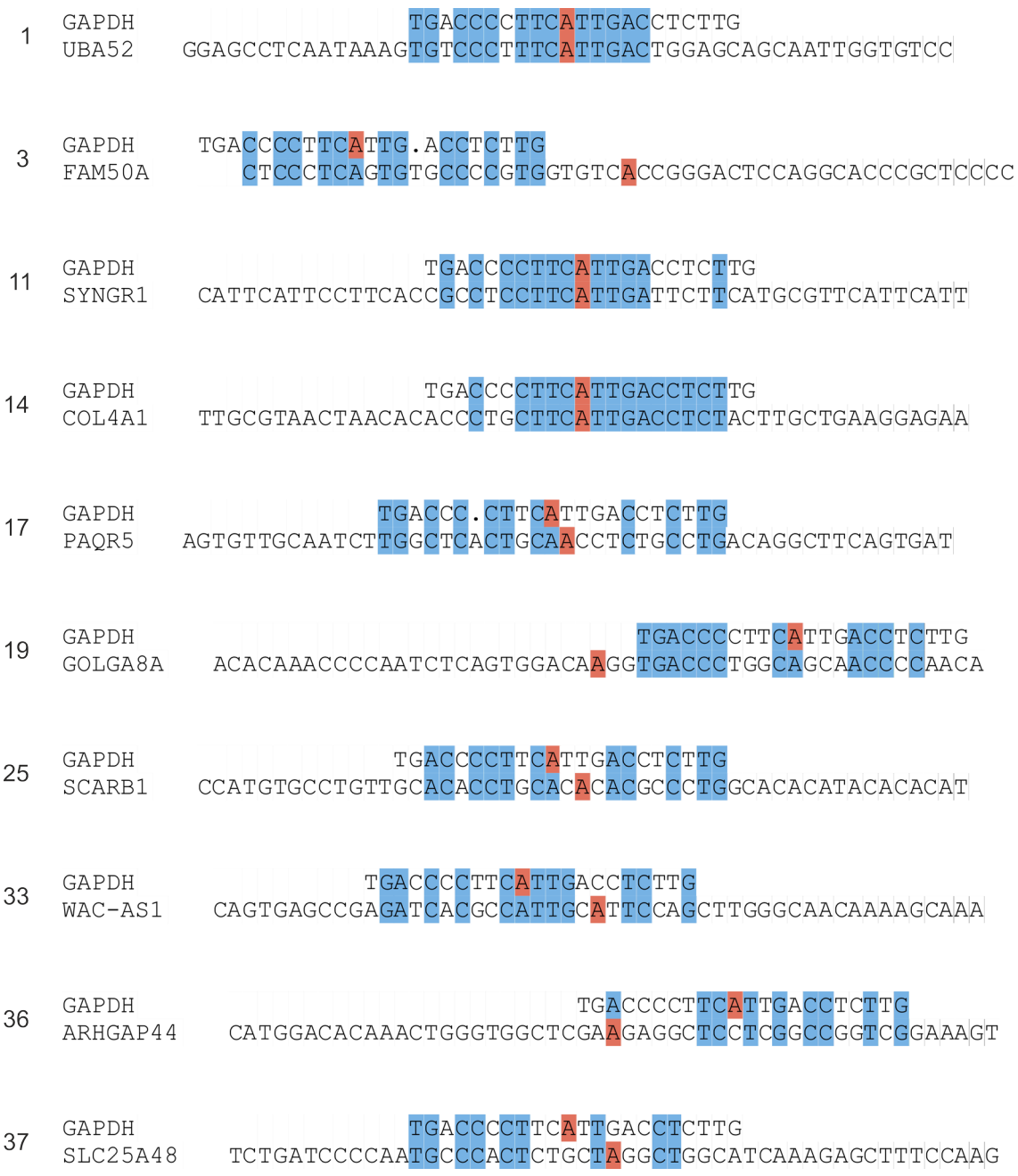


*Figure S17. Alignments of the regions around guideRNA-dependent off-target sites in duo cell line **2** (with entry no. corresponding to Table S7) to the guideRNA-interacting region of the targeted GAPDH transcript. Matching nucleotides are highlighted in blue, the deaminated adenosines in red.*

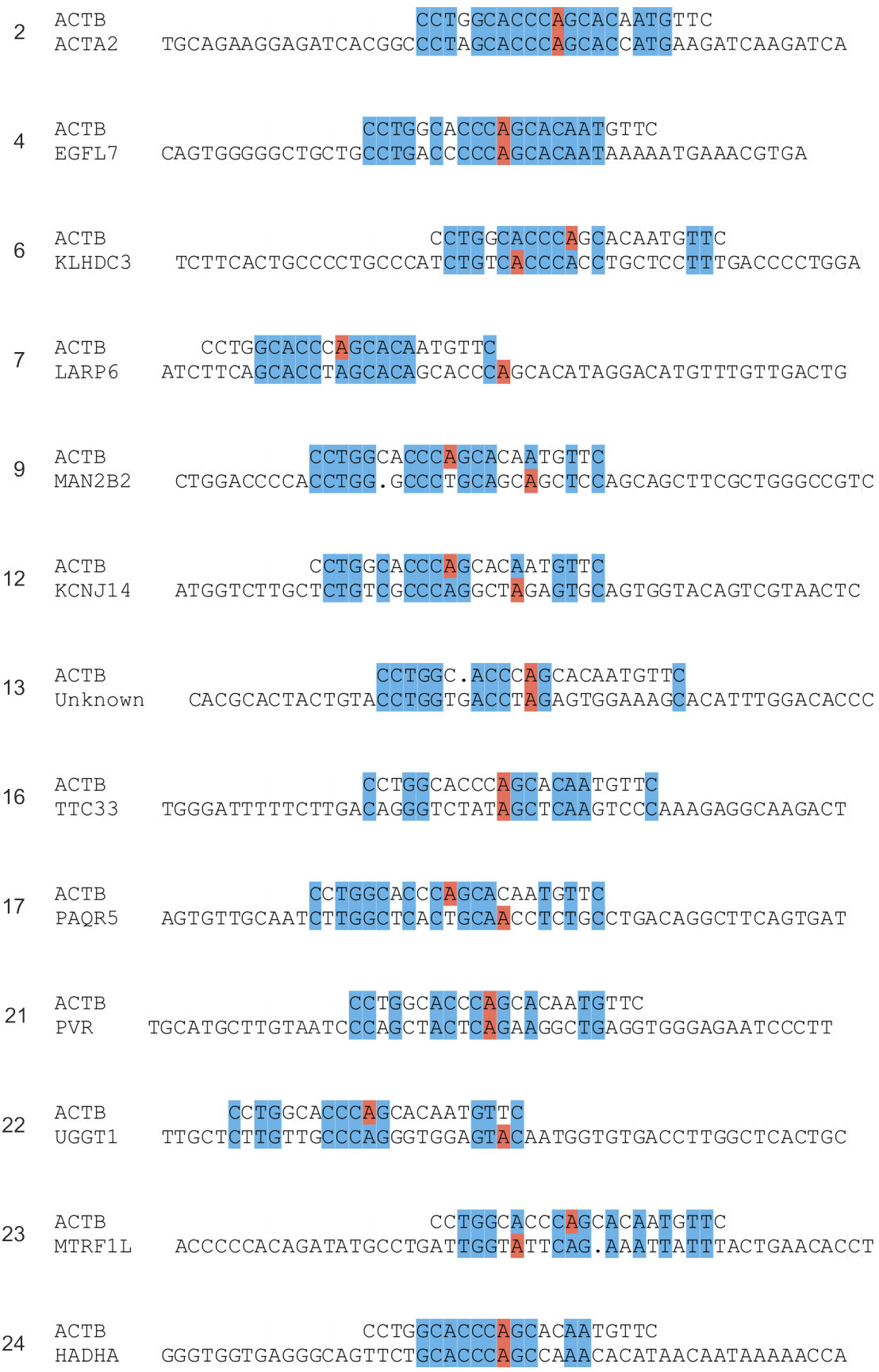


*Figure S18. Alignments of the regions around guideRNA-dependent off-target sites in duo cell line **2** (with entry no. corresponding to Table S7) to the guideRNA-interacting region of the targeted ACTB transcript. Matching nucleotides are highlighted in blue, the deaminated adenosines in red.*

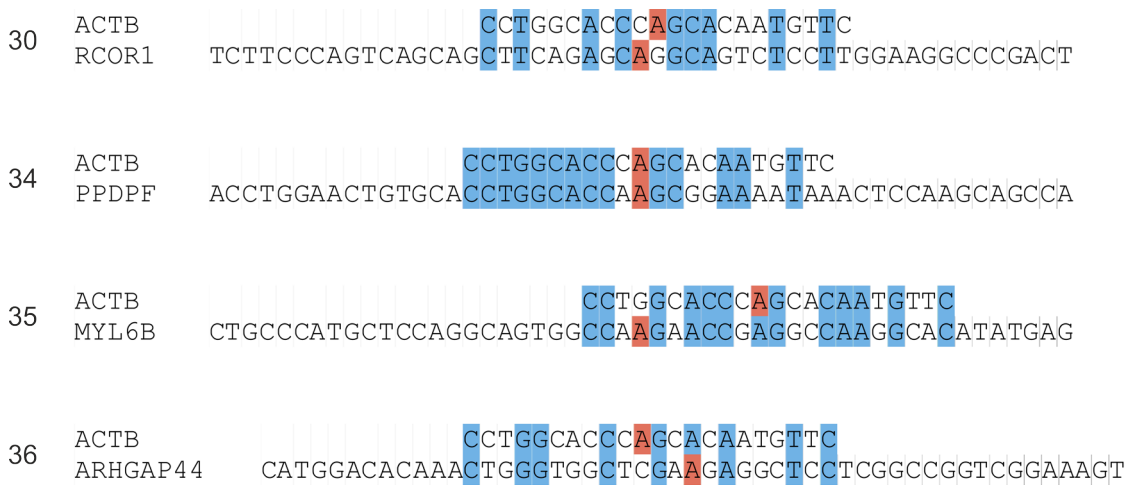


*Figure S18 (continued). Alignments of the regions around guideRNA-dependent off-target sites in duo cell line **2** (with entry no. corresponding to Table S7) to the guideRNA-interacting region of the targeted ACTB transcript. Matching nucleotides are highlighted in blue, the deaminated adenosines in red.*

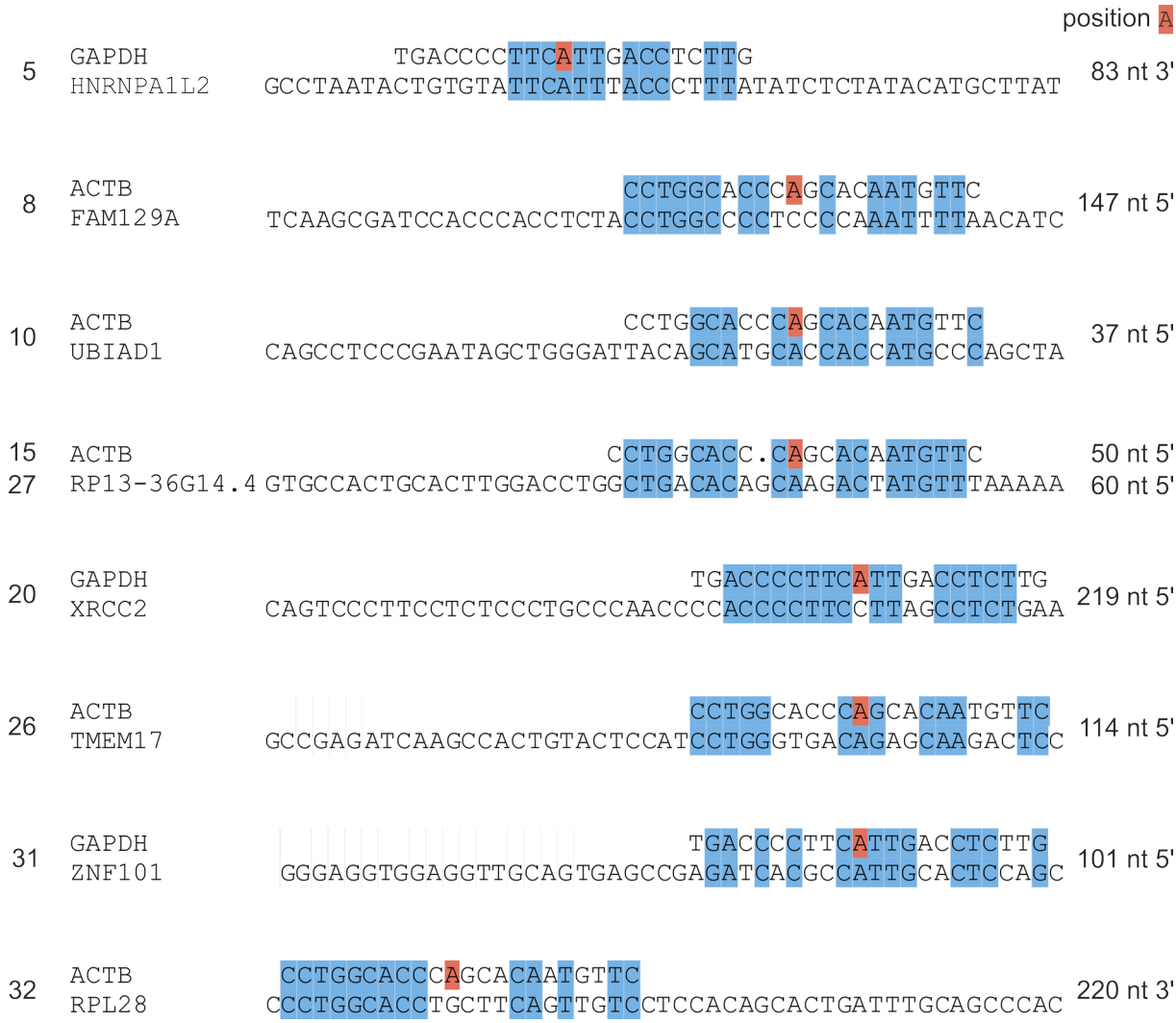


*Figure S19. Alignments of regions around guideRNA-dependent off-target sites in duo cell line **2** in Alu elements (with entry no. corresponding to Table S7) to the guideRNA-interacting regions of either the targeted GAPDH or ACTB transcript. It is likely that the secondary RNA structure within Alu elements*

leads to editable dsRNA once an ADAR is delivered nearby by a guideRNA. The relative position of the off-target site is indicated on the right. Matching nucleotides are highlighted in blue, the targeted adenosine in red.

*Table S8. Bystander editing found by NGS in cell line **2** in GAPDH. Listed are all significantly differently edited sites between duo cell line **2** with guideRNAs versus without guideRNAs within ± 500 bp of the on-target site, sorted by editing difference. In order to spot even minute editing, sites are included independent of editing difference. The on-target site targeted by (halo)$_2$-CAU is shown on the top.*

Entry no.	Distance to on-target / bp	Localization	Editing / % without gRNAs	with gRNAs	difference
On-target	± 0	Missense Mutation	0.0	40.8	40.8
1	+ 142	Missense Mutation	0.2	1.2	1.0
2	+ 187	Missense Mutation	0.0	0.2	0.2
3	+ 175	Missense Mutation	0.1	0.2	0.1
4	+ 141	Missense Mutation	0.0	0.1	0.1

*Table S9. Bystander editing found by NGS in cell line **2** in ACTB. Listed are all significantly differently edited sites between duo cell line **2** with guideRNAs versus without guideRNAs within ± 500 bp of the on-target site, sorted by editing difference. In order to spot even minute editing, sites are included independent of editing difference. The on-target site targeted by (snap)$_2$-CAG is shown on the top.*

Entry no.	Distance to on-target / bp	Localization	Editing / % without gRNAs	with gRNAs	difference
On-target	± 0	Missense Mutation	0.0	51.9	51.9
1	+ 220	Nonstop Mutation	3.7	19.7	16.0
2	+ 221	Missense Mutation	1.0	13.4	12.5
3	– 307	Missense Mutation	0.8	8.5	7.7
4	+ 336	3'UTR	0.0	3.0	3.0
5	+ 224	Missense Mutation	0.0	2.3	2.3
6	+ 425	3'UTR	0.0	1.3	1.3
7	– 325	Missense Mutation	0.1	0.4	0.3
8	– 26	Missense Mutation	0.0	0.1	0.1
9	– 35	Missense Mutation	0.0	0.1	0.1
10	– 365	Missense Mutation	0.0	0.1	0.1

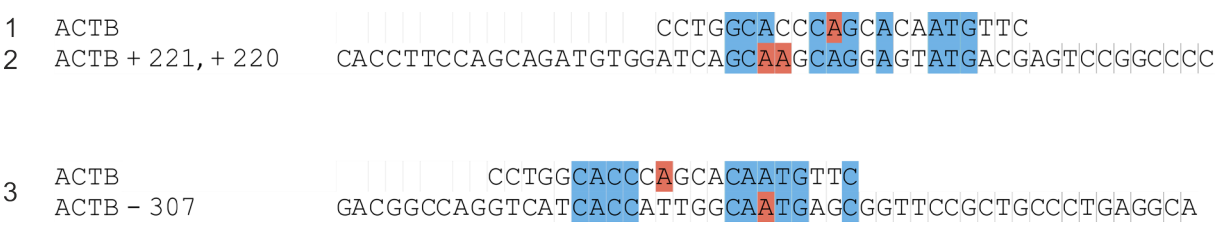


*Figure S20. Alignments of the regions around bystander sites in ACTB in duo cell line **2** with an editing difference ≥ 5 % (with entry no. corresponding to Table S9) to the guideRNA-interacting region of the targeted ACTB transcript. Matching nucleotides are highlighted in blue, the deaminated adenosines in red.*

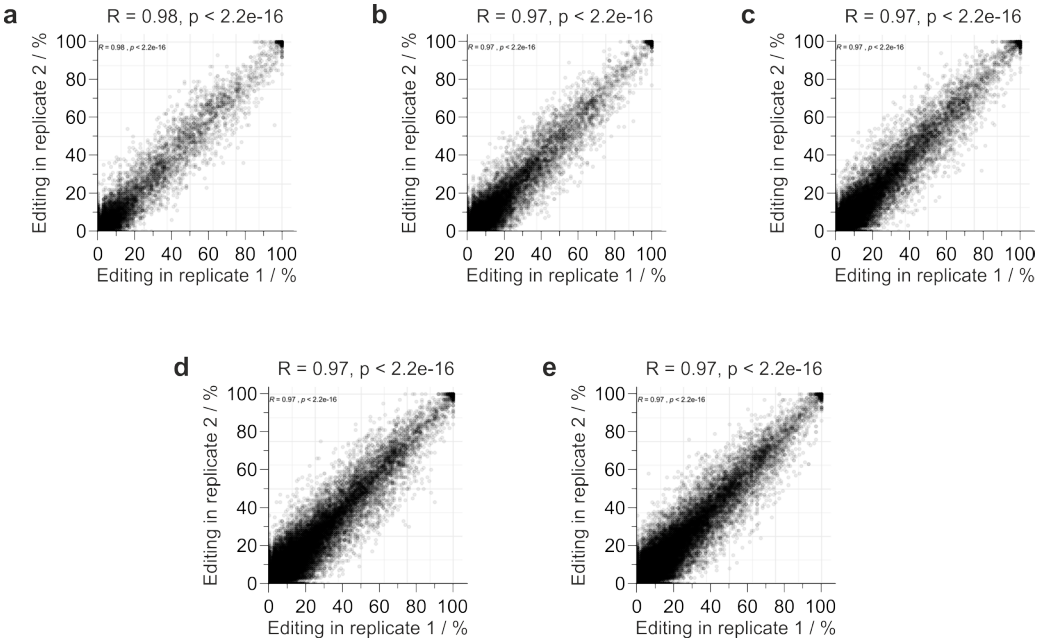


*Figure S21. Scatter plots of all called editing sites in duo cell line **2** in replicate 2 against replicate 1 respectively. **a)** empty 293 Flp-In T-REx. **b)** Cell line expressing SA1Q only after transfection of 5.0 pmol snap-UAG_2.[10] **c)** Cell line expressing SA2Q only after transfection of 5.0 pmol snap-UAG_2.[10] **d)** Cell line **2** (HA1Q + SA2Q) without guideRNA transfection. **e)** Cell line **2** (HA1Q + SA2Q) after transfection of 0.5 pmol $(snap)_2$-CAG and 0.5 pmol $(halo)_2$-CAU.*

HA1Q / APO1S duo cell line

For NGS of duo cell line **9**, again four samples, i.e. a duplicate of an empty transfection and a duplicate of a guideRNA transfection, were prepared analogous to NGS of duo cell line **2**. For the duplicate of the guideRNA transfection, cells were reverse transfected with 2.5 pmol $(halo)_2$-UAU and 2.5 pmol $(snap)_2$-ACC with 0.5 µl Lipofectamine 2000. Cells were harvested 48 h after transfection and subsequently treated as described for NGS of duo cell line **2**. The empty 293 Flp-In T-REx cell line expressing no artificial editing enzyme from Vogel *et al.*[10] (GSM3083474, ctrl_rep1 & GSM3083475, ctrl_rep2) was again applied as negative control. Data analysis is described in the Materials & Methods section.

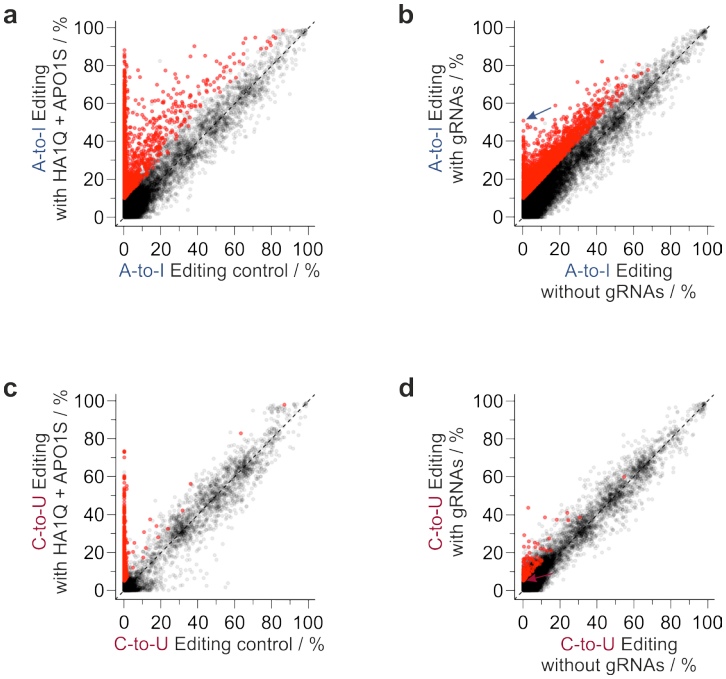


*Figure S22. Scatter plots of off-target analysis of duo cell line **9** (HA1Q + APO1S). Significantly differently edited sites are marked in red. **a)** Total A-to-I off-targets after transfection of 2.5 pmol (halo)$_2$-UAU and 2.5 (snap)$_2$-ACC versus empty 293 Flp-In T-REx.[10] **b)** guideRNA-dependent A-to-I off-targets. The UAU on-target site (ACTB) is marked by a blue arrow. **c)** Total C-to-U off-targets after transfection of 2.5 pmol (halo)$_2$-UAU and 2.5 (snap)$_2$-ACC versus empty 293 Flp-In T-REx.[10] **d)** guideRNA-dependent C-to-U off-targets. The ACC on-target site (GAPDH) is marked by a red arrow.*

*Table S10. Number of significantly differently edited A-to-I sites with editing difference ≥ 25 % found in editing experiments in duo cell line **9** (HA1Q + APO1S) in comparison to a negative control cell line (293 Flp-In T-REx) not expressing any editing enzyme (Total off-targets). The guideRNA-dependent fractions of the total off-targets are shown in the right column.*

	A-to-I (Δ ≥ 25 %)	
	Total off-targets	**gRNA-depend.**
Total number	1621	129
incl. Alu sites	207	8
5'UTR	54	7
Missense mutation	364	49
Nonsense mutation	0	0
Nonstop mutation	18	0
Start codon SNP	1	0
Silent	231	16
3'UTR	715	39
Noncoding	238	19

*Table S 11. Number of significantly differently edited C-to-U sites with editing difference ≥ 10 % and ≥ 25 % found in editing experiments in duo cell line **9** (HA1Q + APO1S) in comparison to a negative control cell line (293 Flp-In T-REx) not expressing any editing enzyme (Total off-targets). The guideRNA-dependent fractions of the total off-targets are shown in the right column respectively.*

	C-to-U (Δ ≥ 10 %)		**C-to-U** (Δ ≥ 25 %)	
	Total off-targets	**gRNA-depend.**	**Total off-targets**	**gRNA-depend.**
Total number	1009	44	129	3
incl. Alu sites	10	1	3	0
5'UTR	21	22	3	0
Missense mutation	7	0	0	0
Nonsense mutation	0	0	0	0
Nonstop mutation	0	0	0	0
Start codon SNP	0	0	0	0
Silent	2	0	0	0
3'UTR	846	33	109	3
Noncoding	133	9	17	0

*Table S12. C-to-U bystander editing found by NGS in cell line **9** in GAPDH. Listed are all significantly differently edited sites between duo cell line **9** with guideRNAs versus without guideRNAs within ± 500 bp of the on-target site, sorted by editing difference. In order to spot even minute editing, sites are included independent of editing difference. The on-target site targeted by $(snap)_2$-ACC is shown on the top, the bystander site at + 472 bp corresponds to one of the off-target sites also observed in Sanger sequencing (see Figure S15b, Table S5 C #2 at + 253, since RNA-seq data is aligned to the human reference genome hg19 including introns).*

Entry no.	**Distance to on-target / bp**	**Localization**	**Editing / %**		
			without gRNAs	**with gRNAs**	**difference**
On-target	± 0	Missense Mutation	0.0	5.1	5.1
1	+ 472	Silent	0.0	6.8	6.8
2	+ 21	Missense Mutation	0.0	0.9	0.9
3	– 3	Missense Mutation	0.0	0.8	0.8
4	– 119	Noncoding	0.1	0.8	0.7
5	+ 25	Silent	0.0	0.7	0.7
6	– 122	Noncoding	0.1	0.8	0.7
7	+ 487	Silent	0.1	0.7	0.6
8	+ 473	Missense Mutation	0.0	0.5	0.5
9	+ 37	Silent	0.0	0.3	0.3
10	– 173	Noncoding	0.1	0.3	0.2
11	+ 49	Silent	0.0	0.2	0.2
12	– 140	Noncoding	0.0	0.2	0.2
13	+ 52	Silent	0.0	0.2	0.2
14	+ 62	Missense Mutation	0.0	0.2	0.2
15	+ 490	Silent	0.0	0.1	0.1
16	– 2	Silent	0.0	0.1	0.1
17	+ 270	Missense Mutation	0.0	0.1	0.1
18	+ 69	Missense Mutation	0.0	0.1	0.1
19	+ 496	Silent	0.0	0.1	0.1
20	+ 256	Silent	0.0	0.1	0.1
21	+ 13	Silent	0.0	0.1	0.1
22	+ 22	Silent	0.0	0.1	0.1

*Table S13. A-to-I bystander editing found by NGS in cell line **9** in ACTB. Listed are all significantly differently edited sites between duo cell line **9** with guideRNAs versus without guideRNAs within ± 500 bp of the on-target site, sorted by editing difference. In order to spot even minute editing, sites are included independent of editing difference. The on-target site targeted by (halo)$_2$-UAU is shown on the top.*

Entry no.	Distance to on-target / bp	Localization	Editing / % without gRNAs	with gRNAs	difference
On-target	± 0	Missense Mutation	0.0	50.8	50.8
22	– 350	Missense Mutation	0.0	0.9	0.9

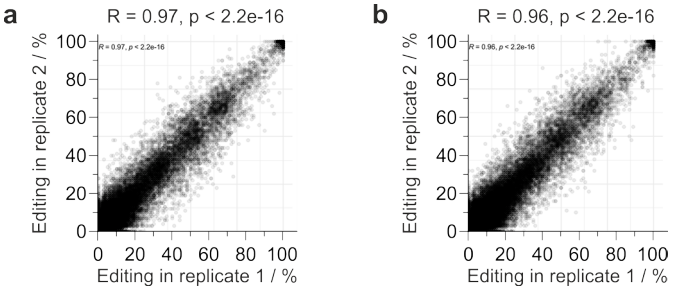


*Figure S23. Scatter plots of all called editing sites in duo cell line **9** in replicate 2 against replicate 1 respectively. **a)** Cell line **9** (HA1Q + APO1S) without guideRNA transfection. **b)** Cell line **9** (HA1Q + APO1S) after transfection of 2.5 pmol (halo)$_2$-UAU and 2.5 (snap)$_2$-ACC.*

Supporting literature

1. Keppler, A. et al. A general method for the covalent labeling of fusion proteins with small molecules *in vivo*. *Nat. Biotechnol.* **21**, 86–89 (2003).
2. Hanswillemenke, A. & Stafforst, T. Protocols for the generation of caged guideRNAs for light-triggered RNA-targeting with SNAP-ADARs. *Methods Enzymol.* **624**, 47–68 (2019).
3. Gautier, A. et al. An Engineered Protein Tag for Multiprotein Labeling in Living Cells. *Chem. Biol.* **15**, 128–136 (2008).
4. Keppler, A. et al. Labeling of fusion proteins of O^6-alkylguanine-DNA alkyltransferase with small molecules in vivo and in vitro. *Methods* **32**, 437–444 (2004).
5. Varga, N. et al. Selective Targeting of Dendritic Cell-Specific Intercellular Adhesion Molecule-3-Grabbing Nonintegrin (DC-SIGN) with Mannose-Based Glycomimetics: Synthesis and Interaction Studies of Bis(benzylamide) Derivatives of a Pseudomannobioside. *Chem. – Eur. J.* **19**, 4786–4797 (2013).
6. Singh, V., Wang, S. & Kool, E.T. Genetically Encoded Multispectral Labeling of Proteins with Polyfluorophores on a DNA Backbone. *J. Am. Chem. Soc.* **135**, 6184–6191 (2013).
7. Los, G.V. et al. HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis. *ACS Chem. Biol.* **3**, 373–382 (2008).
8. Vogel, P. & Stafforst, T. Site-Directed RNA Editing with Antagomir Deaminases — A Tool to Study Protein and RNA Function. *ChemMedChem* **9**, 2021–2025 (2014).
9. Abudayyeh, O.O. et al. A cytosine deaminase for programmable single-base RNA editing. *Science* **365**, 382–386 (2019).
10. Vogel, P. et al. Efficient and precise editing of endogenous transcripts with SNAP-tagged ADARs. *Nat. Methods* **15**, 535–538 (2018).

Appendix

Constructs for single cell lines

Cell line SA1Q: CMV-enhancer – CMV promoter – $TetO_2$ – $SNAP_f$-tag – ADAR1Q – Myc-tag – His-tag – Stop – UAG – bGH

GACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATG
GAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATT
GACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGG
AGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATT
GACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTAC
TTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATG
GGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTG
TTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGG
CGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCCCTATCAGTGATAGAGATCTCCCTA
TCAGTGATAGAGATCGTCGACGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGAGACGCCATCCACGC
TGTTTTGACCTCCATAGAAGACACCGGGACCGATCCAGCCTCCGGACTCTAGCGTTTAAACTTAAGCT
TGGTACCGAGCTCGGATCCATGGACAAAGACTGCGAAATGAAGCGCACCACCCTGGATAGCCCTCTGG
GCAAGCTGGAACTGTCTGGGTGCGAACAGGGCCTGCACCGTATCATCTTCCTGGGCAAAGGAACATCT
GCCGCCGACGCCGTGGAAGTGCCTGCCCCAGCCGCCGTGCTGGGCGGACCAGAGCCACTGATGCAGGC
CACCGCCTGGCTCAACGCCTACTTTCACCAGCCTGAGGCCATCGAGGAGTTCCCTGTGCCAGCCCTGC
ACCACCCAGTGTTCCAGCAGGAGAGCTTTACCCGCCAGGTGCTGTGGAAACTGCTGAAAGTGGTGAAG
TTCGGAGAGGTCATCAGCTACAGCCACCTGGCCGCCCTGGCCGGCAATCCCGCCGCCACCGCCGCCGT
GAAAACCGCCCTGAGCGGAAATCCCGTGCCCATTCTGATCCCCTGCCACCGGGTGGTGCAGGGCGACC
TGGACGTGGGGGGCTACGAGGGCGGGCTCGCCGTGAAAGAGTGGCTGCTGGCCCACGAGGGCCACAGA
CTGGGCAAGCCTGGGCTGGGTCCTGCAGGCGGAGGCGCGCCAGGGTCTGGCGGCGGCAGTAAGGCAGA
ACGCATGGGTTTCACAGAGGTAACCCCAGTGACAGGGGCCAGTCTCAGAAGAACTATGCTCCTCCTCT
CAAGGTCCCCAGAAGCACAGCCAAAGACACTCCCTCTCACTGGCAGCACCTTCCATGACCAGATAGCC
ATGCTGAGCCACCGGTGCTTCAACACTCTGACTAACAGCTTCCAGCCCTCCTTGCTCGGCCGCAAGAT
TCTGGCCGCCATCATTATGAAAAAAGACTCTGAGGACATGGGTGTCGTCGTCAGCTTGGGAACAGGGA
ATCGCTGTGTAAAAGGAGATTCTCTCAGCCTAAAAGGAGAAACTGTCAATGACTGCCATGCAGAAATA
ATCTCCCGGAGAGGCTTCATCAGGTTTCTCTACAGTGAGTTAATGAAATACAACTCCCAGACTGCGAA
GGATAGTATATTTGAACCTGCTAAGGGAGGAGAAAAGCTCCAAATAAAAAAGACTGTGTCATTCCATC
TGTATATCAGCACTGCTCCGTGTGGAGATGGCGCCCTCTTTGACAAGTCCTGCAGCGACCGTGCTATG
GAAAGCACAGAATCCCGCCACTACCCTGTCTTCGAGAATCCCAAACAAGGAAAGCTCCGCACCAAGGT
GGAGAACGGACAAGGCACAATCCCTGTGGAATCCAGTGACATTGTGCCTACGTGGGATGGCATTCGGC
TCGGGGAGAGACTCCGTACCATGTCCTGTAGTGACAAAATCCTACGCTGGAACGTGCTGGGCCTGCAA
GGGGCACTGTTGACCCACTTCCTGCAGCCCATTTATCTCAAATCTGTCACATTGGGTTACCTTTTCAG
CCAAGGGCATCTGACCCGTGCTATTTGCTGTCGTGTGACAAGAGATGGGAGTGCATTTGAGGATGGAC
TACGACATCCCTTTATTGTCAACCACCCCAAGGTTGGCAGAGTCAGCATATATGATTCCAAAAGGCAA
TCCGGGAAGACTAAGGAGACAAGCGTCAACTGGTGTCTGGCTGATGGCTATGACCTGGAGATCCTGGA
CGGTACCAGAGGCACTGTGGATGGGCCACGGAATGAATTGTCCCGGGTCTCCAAAAAGAACATTTTTC
TTCTATTTAAGAAGCTCTGCTCCTTCCGTTACCGCAGGGATCTACTGAGACTCTCCTATGGTGAGGCC
AAGAAAGCTGCCCGTGACTACGAGACGGCCAAGAACTACTTCAAAAAAGGCCTGAAGGATATGGGCTA
TGGGAACTGGATTAGCAAACCCCAGGAGGAAAAGAACTTTTATCTCTGCCCAGTATCTAGAGGGCCCT
TCGAACAAAAACTCATCTCAGAAGAGGATCTGAATATGCATACCGGTCATCATCACCATCACCATTGA
CTGCCTGTTCCGTAGCCGACACGCGGCCGCTCGAGTCTAGAGGGCCCGTTTAAACCCGCTGATCAGCC
TCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGA
AGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTC
ATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCAT
GCTGGGGATGCGGTGGGCTCTATGG

Cell line CA1Q: CMV-enhancer – CMV promoter – $TetO_2$ – $CLIP_f$-tag – ADAR1Q – Myc-tag – His-tag – Stop – UAG – bGH

GACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATG
GAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATT
GACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGG
AGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATT
GACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTAC
TTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATG
GGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTG
TTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGG
CGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCCCTATCAGTGATAGAGATCTCCCTA
TCAGTGATAGAGATCGTCGACGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGAGACGCCATCCACGC
TGTTTTGACCTCCATAGAAGACACCGGGACCGATCCAGCCTCCGGACTCTAGCGTTTAAACTTAAGCT
TGGTACCGAGCTCGGATCCATGGACAAAGACTGCGAAATGAAGCGCACCACCCTGGATAGCCCTCTGG
GCAAGCTGGAACTGTCTGGGTGCGAACAGGGCCTGCACCGTATCATCTTCCTGGGCAAAGGAACATCT
GCCGCCGACGCCGTGGAAGTGCCTGCCCCAGCCGCCGTGCTGGGCGGACCAGAGCCACTGATCCAGGC
CACCGCCTGGCTCAACGCCTACTTTCACCAGCCTGAGGCCATCGAGGAGTTCCCTGTGCCAGCCCTGC
ACCACCCAGTGTTCCAGCAGGAGAGCTTTACCCGCCAGGTGCTGTGGAAACTGCTGAAAGTGGTGAAG
TTCGGAGAGGTCATCAGCGAGAGCCACCTGGCCGCCCTGGTGGGCAATCCCGCCGCCACCGCCGCCGT
GAACACCGCCCTGGACGGAAATCCCGTGCCCATTCTGATCCCCTGCCACCGGGTGGTGCAGGGCGACA
GCGACGTGGGGCCCTACCTGGGCGGGCTCGCCGTGAAAGAGTGGCTGCTGGCCCACGAGGGCCACAGA
CTGGGCAAGCCTGGGCTGGGTCCTGCAGGCGGAGGCGCGCCAGGGTCTGGCGGCGGCAGTAAGGCAGA
ACGCATGGGTTTCACAGAGGTAACCCCAGTGACAGGGGCCAGTCTCAGAAGAACTATGCTCCTCCTCT
CAAGGTCCCCAGAAGCACAGCCAAAGACACTCCCTCTCACTGGCAGCACCTTCCATGACCAGATAGCC
ATGCTGAGCCACCGGTGCTTCAACACTCTGACTAACAGCTTCCAGCCCTCCTTGCTCGGCCGCAAGAT
TCTGGCCGCCATCATTATGAAAAAAGACTCTGAGGACATGGGTGTCGTCGTCAGCTTGGGAACAGGGA
ATCGCTGTGTAAAAGGAGATTCTCTCAGCCTAAAAGGAGAAACTGTCAATGACTGCCATGCAGAAATA
ATCTCCCGGAGAGGCTTCATCAGGTTTCTCTACAGTGAGTTAATGAAATACAACTCCCAGACTGCGAA
GGATAGTATATTTGAACCTGCTAAGGGAGGAGAAAAGCTCCAAATAAAAAAGACTGTGTCATTCCATC
TGTATATCAGCACTGCTCCGTGTGGAGATGGCGCCCTCTTTGACAAGTCCTGCAGCGACCGTGCTATG
GAAAGCACAGAATCCCGCCACTACCCTGTCTTCGAGAATCCCAAACAAGGAAAGCTCCGCACCAAGGT
GGAGAACGGACAAGGCACAATCCCTGTGGAATCCAGTGACATTGTGCCTACGTGGGATGGCATTCGGC
TCGGGGAGAGACTCCGTACCATGTCCTGTAGTGACAAAATCCTACGCTGGAACGTGCTGGGCCTGCAA
GGGGCACTGTTGACCCACTTCCTGCAGCCCATTTATCTCAAATCTGTCACATTGGGTTACCTTTTCAG
CCAAGGGCATCTGACCCGTGCTATTTGCTGTCGTGTGACAAGAGATGGGAGTGCATTTGAGGATGGAC
TACGACATCCCTTTATTGTCAACCACCCCAAGGTTGGCAGAGTCAGCATATATGATTCCAAAAGGCAA
TCCGGGAAGACTAAGGAGACAAGCGTCAACTGGTGTCTGGCTGATGGCTATGACCTGGAGATCCTGGA
CGGTACCAGAGGCACTGTGGATGGGCCACGGAATGAATTGTCCCGGGTCTCCAAAAAGAACATTTTTC
TTCTATTTAAGAAGCTCTGCTCCTTCCGTTACCGCAGGGATCTACTGAGACTCTCCTATGGTGAGGCC
AAGAAAGCTGCCCGTGACTACGAGACGGCCAAGAACTACTTCAAAAAAGGCCTGAAGGATATGGGCTA
TGGGAACTGGATTAGCAAACCCCAGGAGGAAAAGAACTTTTATCTCTGCCCAGTATCTAGAGGGCCCT
TCGAACAAAAACTCATCTCAGAAGAGGATCTGAATATGCATACCGGTCATCATCACCATCACCATTGA
CTGCCTGTTCCGTAGCCGACACGCGGCCGCTCGAGTCTAGAGGGCCCGTTTAAACCCGCTGATCAGCC
TCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGA
AGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTC
ATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCAT
GCTGGGGATGCGGTGGGCTCTATGG

Cell line HA1Q: CMV-enhancer – CMV promoter – $TetO_2$ – HaloTag – ADAR1Q – Myc-tag – His-tag – Stop – UAG – bGH

GACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATG
GAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATT
GACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGG
AGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATT
GACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTAC
TTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATG
GGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTG
TTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGG
CGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCCCTATCAGTGATAGAGATCTCCCTA
TCAGTGATAGAGATCGTCGACGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGAGACGCCATCCACGC
TGTTTTGACCTCCATAGAAGACACCGGGACCGATCCAGCCTCCGGACTCTAGCGTTTAAACTTAAGCT
TGGTACCGAGCTCGGATCCATGGCAGAAATCGGTACTGGCTTTCCATTCGACCCCCATTATGTGGAAG
TCCTGGGCGAGCGCATGCACTACGTCGATGTTGGTCCGCGCGATGGCACCCCTGTGCTGTTCCTGCAC
GGTAACCCGACCTCCTCCTACGTGTGGCGCAACATCATCCCGCATGTTGCACCGACCCATCGCTGCAT
TGCTCCAGACCTGATCGGTATGGGCAAATCCGACAAACCAGACCTGGGTTATTTCTTCGACGACCACG
TCCGCTTCATGGATGCCTTCATCGAAGCCCTGGGTCTGGAAGAGGTCGTCCTGGTCATTCACGACTGG
GGCTCCGCTCTGGGTTTCCACTGGGCCAAGCGCAATCCAGAGCGCGTCAAAGGTATTGCATTTATGGA
GTTCATCCGCCCTATCCCGACCTGGGACGAATGGCCAGAATTTGCCCGCGAGACCTTCCAGGCCTTCC
GCACCACCGACGTCGGCCGCAAGCTGATCATCGATCAGAACGTTTTTATCGAGGGTACGCTGCCGATG
GGTGTCGTCCGCCCGCTGACTGAAGTCGAGATGGACCATTACCGCGAGCCGTTCCTGAATCCTGTTGA
CCGCGAGCCACTGTGGCGCTTCCCAAACGAGCTGCCAATCGCCGGTGAGCCAGCGAACATCGTCGCGC
TGGTCGAAGAATACATGGACTGGCTGCACCAGTCCCCTGTCCCGAAGCTGCTGTTCTGGGGCACCCCA
GGCGTTCTGATCCCACCGGCCGAAGCCGCTCGCCTGGCCAAAAGCCTGCCTAACTGCAAGGCTGTGGA
CATCGGCCCGGGTCTGAATCTGCTGCAAGAAGACAACCCGGACCTGATCGGCAGCGAGATCGCGCGCT
GGCTGTCGACGCTCGAGAAGCCAACCCCTGCAGGCGGAGGCGCGCCAGGGTCTGGCGGCGGCAGTAAG
GCAGAACGCATGGGTTTCACAGAGGTAACCCCAGTGACAGGGGCCAGTCTCAGAAGAACTATGCTCCT
CCTCTCAAGGTCCCCAGAAGCACAGCCAAAGACACTCCCTCTCACTGGCAGCACCTTCCATGACCAGA
TAGCCATGCTGAGCCACCGGTGCTTCAACACTCTGACTAACAGCTTCCAGCCCTCCTTGCTCGGCCGC
AAGATTCTGGCCGCCATCATTATGAAAAAAGACTCTGAGGACATGGGTGTCGTCGTCAGCTTGGGAAC
AGGGAATCGCTGTGTAAAAGGAGATTCTCTCAGCCTAAAAGGAGAAACTGTCAATGACTGCCATGCAG
AAATAATCTCCCGGAGAGGCTTCATCAGGTTTCTCTACAGTGAGTTAATGAAATACAACTCCCAGACT
GCGAAGGATAGTATATTTGAACCTGCTAAGGGAGGAGAAAAGCTCCAAATAAAAAAGACTGTGTCATT
CCATCTGTATATCAGCACTGCTCCGTGTGGAGATGGCGCCCTCTTTGACAAGTCCTGCAGCGACCGTG
CTATGGAAAGCACAGAATCCCGCCACTACCCTGTCTTCGAGAATCCCAAACAAGGAAAGCTCCGCACC
AAGGTGGAGAACGGACAAGGCACAATCCCTGTGGAATCCAGTGACATTGTGCCTACGTGGGATGGCAT
TCGGCTCGGGGAGAGACTCCGTACCATGTCCTGTAGTGACAAAATCCTACGCTGGAACGTGCTGGGCC
TGCAAGGGGCACTGTTGACCCACTTCCTGCAGCCCATTTATCTCAAATCTGTCACATTGGGTTACCTT
TTCAGCCAAGGGCATCTGACCCGTGCTATTTGCTGTCGTGTGACAAGAGATGGGAGTGCATTTGAGGA
TGGACTACGACATCCCTTTATTGTCAACCACCCCAAGGTTGGCAGAGTCAGCATATATGATTCCAAAA
GGCAATCCGGGAAGACTAAGGAGACAAGCGTCAACTGGTGTCTGGCTGATGGCTATGACCTGGAGATC
CTGGACGGTACCAGAGGCACTGTGGATGGGCCACGGAATGAATTGTCCCGGGTCTCCAAAAAGAACAT
TTTTCTTCTATTTAAGAAGCTCTGCTCCTTCCGTTACCGCAGGGATCTACTGAGACTCTCCTATGGTG
AGGCCAAGAAAGCTGCCCGTGACTACGAGACGGCCAAGAACTACTTCAAAAAAGGCCTGAAGGATATG
GGCTATGGGAACTGGATTAGCAAACCCCAGGAGGAAAAGAACTTTTATCTCTGCCCAGTATCTAGAGG
GCCCTTCGAACAAAAACTCATCTCAGAAGAGGATCTGAATATGCATACCGGTCATCATCACCATCACC
ATTGACTGCCTGTTCCGTAGCCGACACGCGGCCGCTCGAGTCTAGAGGGCCCGTTTAAACCCGCTGAT
CAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACC
CTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAG
GTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCA
GGCATGCTGGGGATGCGGTGGGCTCTATGG

Constructs for HA1Q / SA2Q duo cell lines 1 – 5

Cell line **1**: CMV-enhancer – CMV promoter – $TetO_2$ – $SNAP_f$-tag – ADAR2Q – bGH – CMV-enhancer – CMV promoter – $TetO_2$ – HaloTag – ADAR1Q – bGH

GACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATG
GAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATT
GACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGG
AGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATT
GACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTAC
TTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATG
GGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTG
TTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGG
CGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCCCTATCAGTGATAGAGATCTCCCTA
TCAGTGATAGAGATCGTCGACGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGAGACGCCATCCACGC
TGTTTTGACCTCCATAGAAGACACCGGGACCGATCCAGCCTCCGGACTCTAGCGTTTAAACTTAAGCT
TGGTACCGAGCTCGGATCCCCACCATGGACAAAGACTGCGAAATGAAGCGCACCACCCTGGATAGCCC
TCTGGGCAAGCTGGAACTGTCTGGGTGCGAACAGGGCCTGCACCGTATCATCTTCCTGGGCAAAGGAA
CATCTGCCGCCGACGCCGTGGAAGTGCCTGCCCCAGCCGCCGTGCTGGGCGGACCAGAGCCACTGATG
CAGGCCACCGCCTGGCTCAACGCCTACTTTCACCAGCCTGAGGCCATCGAGGAGTTCCCTGTGCCAGC
CCTGCACCACCCAGTGTTCCAGCAGGAGAGCTTTACCCGCCAGGTGCTGTGGAAACTGCTGAAAGTGG
TGAAGTTCGGAGAGGTCATCAGCTACAGCCACCTGGCCGCCCTGGCCGGCAATCCCGCCGCCACCGCC
GCCGTGAAAACCGCCCTGAGCGGAAATCCCGTGCCCATTCTGATCCCCTGCCACCGGGTGGTGCAGGG
CGACCTGGACGTGGGGGGCTACGAGGGCGGGCTCGCCGTGAAAGAGTGGCTGCTGGCCCACGAGGGCC
ACAGACTGGGCAAGCCTGGGCTGGGTCCTGCAGGCGGAGGCGCGCCAGGGTCTGGCGGCGGCAGTAAG
AAGCTTGCCAAGGCCCGGGCTGCGCAGTCTGCCCTGGCCGCCATTTTTAACTTGCACTTGGATCAGAC
GCCATCTCGCCAGCCTATTCCCAGTGAGGGTCTTCAGCTGCATTTACCGCAGGTTTTAGCTGACGCTG
TCTCACGCCTGGTCCTGGGTAAGTTTGGTGACCTGACCGACAACTTCTCCTCCCCTCACGCTCGCAGA
AAAGTGCTGGCTGGAGTCGTCATGACAACAGGCACAGATGTTAAAGATGCCAAGGTGATAAGTGTTTC
TACAGGAACAAAATGTATTAATGGTGAATACATGAGTGATCGTGGCCTTGCATTAAATGACTGCCATG
CAGAAATAATATCTCGGAGATCCTTGCTCAGATTTCTTTATACACAACTTGAGCTTTACTTAAATAAC
AAAGATGATCAAAAAAGATCCATCTTTCAGAAATCAGAGCGAGGGGGGTTTAGGCTGAAGGAGAATGT
CCAGTTTCATCTGTACATCAGCACCTCTCCCTGTGGAGATGCCAGAATCTTCTCACCACATGAGCCAA
TCCTGGAAGAACCAGCAGATAGACACCCAAATCGTAAAGCAAGAGGACAGCTACGGACCAAAATAGAG
TCTGGTCAGGGGACGATTCCAGTGCGCTCCAATGCGAGCATCCAAACGTGGGACGGGGTGCTGCAAGG
GGAGCGGCTGCTCACCATGTCCTGCAGTGACAAGATTGCACGCTGGAACGTGGTGGGCATCCAGGGTT
CCCTGCTCAGCATTTTCGTGGAGCCCATTTACTTCTCGAGCATCATCCTGGGCAGCCTTTACCACGGG
GACCACCTTTCCAGGGCCATGTACCAGCGGATCTCCAACATAGAGGACCTGCCACCTCTCTACACCCT
CAACAAGCCTTTGCTCAGTGGCATCAGCAATGCAGAAGCACGGCAGCCAGGGAAGGCCCCCAACTTCA
GTGTCAACTGGACGGTAGGCGACTCCGCTATTGAGGTCATCAACGCCACGACTGGGAAGGATGAGCTG
GGCCGCGCGTCCCGCCTGTGTAAGCACGCGTTGTACTGTCGCTGGATGCGTGTGCACGGCAAGGTTCC
CTCCCACTTACTACGCTCCAAGATTACCAAACCCAACGTGTACCATGAGTCCAAGCTGGCGGCAAAGG
AGTACCAGGCCGCCAAGGCGCGTCTGTTCACAGCCTTCATCAAGGCGGGGCTGGGGGCCTGGGTGGAG
AAGCCCACCGAGCAGGACCAGTTCTCACTCACGCCCTGAGGGCCCATGTACGATTTAATTATGCGGAC
GTGATGAGCGAAGTACGATCCCACGACCGAGGCCCGTTTAAACCCGCTGATCAGCCTCGACTGTGCCT
TCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCC
CACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGG
GGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCG
GTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTATCCCCACGCGCCCTG
TAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCC
TAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGTCGATGTACGGGCCAG
ATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATA
GCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGAC

CCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACG
TCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTA
CGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGG
GACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCA
GTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCA
ATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTG
ACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCCCTATCAGTGATAG
AGATCTCCCTATCAGTGATAGAGAGCGTGCATAGGGAACATCCACCACTTTAGTGAATTGTAGCACGG
CTTCAGAAGCGGCCGCCCACCATGGCAGAAATCGGTACTGGCTTTCCATTCGACCCCCATTATGTGGA
AGTCCTGGGCGAGCGCATGCACTACGTCGATGTTGGTCCGCGCGATGGCACCCCTGTGCTGTTCCTGC
ACGGTAACCCGACCTCCTCCTACGTGTGGCGCAACATCATCCCGCATGTTGCACCGACCCATCGCTGC
ATTGCTCCAGACCTGATCGGTATGGGCAAATCCGACAAACCAGACCTGGGTTATTTCTTCGACGACCA
CGTCCGCTTCATGGATGCCTTCATCGAAGCCCTGGGTCTGGAAGAGGTCGTCCTGGTCATTCACGACT
GGGGCTCCGCTCTGGGTTTCCACTGGGCCAAGCGCAATCCAGAGCGCGTCAAAGGTATTGCATTTATG
GAGTTCATCCGCCCTATCCCGACCTGGGACGAATGGCCAGAATTTGCCCGCGAGACCTTCCAGGCCTT
CCGCACCACCGACGTCGGCCGCAAGCTGATCATTGATCAGAACGTTTTTATCGAGGGTACGCTGCCGA
TGGGTGTCGTCCGCCCGCTGACTGAAGTCGAGATGGACCATTACCGCGAGCCGTTCCTGAATCCTGTT
GACCGCGAGCCACTGTGGCGCTTCCCAAACGAGCTGCCAATCGCCGGTGAGCCAGCGAACATCGTCGC
GCTGGTCGAAGAATACATGGACTGGCTGCACCAGTCCCCTGTCCCGAAGCTGCTGTTCTGGGGCACCC
CAGGCGTTCTGATCCCACCGGCCGAAGCCGCTCGCCTGGCCAAAAGCCTGCCTAACTGCAAGGCTGTG
GACATCGGCCCGGGTCTGAATCTGCTGCAAGAAGACAACCCGGACCTGATCGGCAGCGAGATCGCGCG
CTGGCTGTCGACGCTCGAGATTTCCGGCCCTGCAGGCGGAGGCGCGCCAGGGTCTGGCGGCGGCAGTA
AGGCAGAACGCATGGGTTTCACAGAGGTAACCCCAGTGACAGGGGCCAGTCTCAGAAGAACTATGCTC
CTCCTCTCAAGGTCCCCAGAAGCACAGCCAAAGACACTCCCTCTCACTGGCAGCACCTTCCATGACCA
GATAGCCATGCTGAGCCACCGGTGCTTCAACACTCTGACTAACAGCTTCCAGCCCTCCTTGCTCGGCC
GCAAGATTCTGGCCGCCATCATTATGAAAAAAGACTCTGAGGACATGGGTGTCGTCGTCAGCTTGGGA
ACAGGGAATCGCTGTGTAAAAGGAGATTCTCTCAGCCTAAAAGGAGAAACTGTCAATGACTGCCATGC
AGAAATAATCTCCCGGAGAGGCTTCATCAGGTTTCTCTACAGTGAGTTAATGAAATACAACTCCCAGA
CTGCGAAGGATAGTATATTTGAACCTGCTAAGGGAGGAGAAAAGCTCCAAATAAAAAAGACTGTGTCA
TTCCATCTGTATATCAGCACTGCTCCGTGTGGAGATGGCGCCCTCTTTGACAAGTCCTGCAGCGACCG
TGCTATGGAAAGCACAGAATCCCGCCACTACCCTGTCTTCGAGAATCCCAAACAAGGAAAGCTCCGCA
CCAAGGTGGAGAACGGACAAGGCACAATCCCTGTGGAATCCAGTGACATTGTGCCTACGTGGGATGGC
ATTCGGCTCGGGGAGAGACTCCGTACCATGTCCTGTAGTGACAAAATCCTACGCTGGAACGTGCTGGG
CCTGCAAGGGGCACTGTTGACCCACTTCCTGCAGCCCATTTATCTCAAATCTGTCACATTGGGTTACC
TTTTCAGCCAAGGGCATCTGACCCGTGCTATTTGCTGTCGTGTGACAAGAGATGGGAGTGCATTTGAG
GATGGACTACGACATCCCTTTATTGTCAACCACCCCAAGGTTGGCAGAGTCAGCATATATGATTCCAA
AAGGCAATCCGGGAAGACTAAGGAGACAAGCGTCAACTGGTGTCTGGCTGATGGCTATGACCTGGAGA
TCCTGGACGGTACCAGAGGCACTGTGGATGGGCCACGGAATGAATTGTCCCGGGTCTCCAAAAAGAAC
ATTTTTCTTCTATTTAAGAAGCTCTGCTCCTTCCGTTACCGCAGGGATCTACTGAGACTCTCCTATGG
TGAGGCCAAGAAAGCTGCCCGTGACTACGAGACGGCCAAGAACTACTTCAAAAAAGGCCTGAAGGATA
TGGGCTATGGGAACTGGATTAGCAAACCCCAGGAGGAAAAGAACTTTTATCTCTGCCCAGTATGAATC
GATATTTTCAGATATCGTGTTAGTAGGGTTGCACCGACGCGCATGTGGATTAGTGCTGTGCCTTCTAG
TTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTG
TCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGT
GGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGG
CTCTATGG

Cell line **2**: CMV-enhancer – CMV promoter – $TetO_2$ – HaloTag – ADAR1Q – bGH – CMV-enhancer – CMV promoter – $TetO_2$ – $SNAP_f$-tag– ADAR2Q – bGH

GACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATG
GAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATT
GACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGG
AGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATT
GACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTAC
TTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATG
GGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTG
TTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGG
CGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCCCTATCAGTGATAGAGATCTCCCTA
TCAGTGATAGAGATCGTCGACGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGAGACGCCATCCACGC
TGTTTTGACCTCCATAGAAGACACCGGGACCGATCCAGCCTCCGGACTCTAGCGTTTAAACTTAAGCT
TGGTACCGAGCTCGGATCCCCACCATGGCAGAAATCGGTACTGGCTTTCCATTCGACCCCCATTATGT
GGAAGTCCTGGGCGAGCGCATGCACTACGTCGATGTTGGTCCGCGCGATGGCACCCCTGTGCTGTTCC
TGCACGGTAACCCGACCTCCTCCTACGTGTGGCGCAACATCATCCCGCATGTTGCACCGACCCATCGC
TGCATTGCTCCAGACCTGATCGGTATGGGCAAATCCGACAAACCAGACCTGGGTTATTTCTTCGACGA
CCACGTCCGCTTCATGGATGCCTTCATCGAAGCCCTGGGTCTGGAAGAGGTCGTCCTGGTCATTCACG
ACTGGGGCTCCGCTCTGGGTTTCCACTGGGCCAAGCGCAATCCAGAGCGCGTCAAAGGTATTGCATTT
ATGGAGTTCATCCGCCCTATCCCGACCTGGGACGAATGGCCAGAATTTGCCCGCGAGACCTTCCAGGC
CTTCCGCACCACCGACGTCGGCCGCAAGCTGATCATTGATCAGAACGTTTTTATCGAGGGTACGCTGC
CGATGGGTGTCGTCCGCCCGCTGACTGAAGTCGAGATGGACCATTACCGCGAGCCGTTCCTGAATCCT
GTTGACCGCGAGCCACTGTGGCGCTTCCCAAACGAGCTGCCAATCGCCGGTGAGCCAGCGAACATCGT
CGCGCTGGTCGAAGAATACATGGACTGGCTGCACCAGTCCCCTGTCCCGAAGCTGCTGTTCTGGGGCA
CCCCAGGCGTTCTGATCCCACCGGCCGAAGCCGCTCGCCTGGCCAAAAGCCTGCCTAACTGCAAGGCT
GTGGACATCGGCCCGGGTCTGAATCTGCTGCAAGAAGACAACCCGGACCTGATCGGCAGCGAGATCGC
GCGCTGGCTGTCGACGCTCGAGATTTCCGGCCCTGCAGGCGGAGGCGCGCCAGGGTCTGGCGGCGGCA
GTAAGGCAGAACGCATGGGTTTCACAGAGGTAACCCCAGTGACAGGGGCCAGTCTCAGAAGAACTATG
CTCCTCCTCTCAAGGTCCCCAGAAGCACAGCCAAAGACACTCCCTCTCACTGGCAGCACCTTCCATGA
CCAGATAGCCATGCTGAGCCACCGGTGCTTCAACACTCTGACTAACAGCTTCCAGCCCTCCTTGCTCG
GCCGCAAGATTCTGGCCGCCATCATTATGAAAAAAGACTCTGAGGACATGGGTGTCGTCGTCAGCTTG
GGAACAGGGAATCGCTGTGTAAAAGGAGATTCTCTCAGCCTAAAAGGAGAAACTGTCAATGACTGCCA
TGCAGAAATAATCTCCCGGAGAGGCTTCATCAGGTTTCTCTACAGTGAGTTAATGAAATACAACTCCC
AGACTGCGAAGGATAGTATATTTGAACCTGCTAAGGGAGGAGAAAAGCTCCAAATAAAAAAGACTGTG
TCATTCCATCTGTATATCAGCACTGCTCCGTGTGGAGATGGCGCCCTCTTTGACAAGTCCTGCAGCGA
CCGTGCTATGGAAAGCACAGAATCCCGCCACTACCCTGTCTTCGAGAATCCCAAACAAGGAAAGCTCC
GCACCAAGGTGGAGAACGGACAAGGCACAATCCCTGTGGAATCCAGTGACATTGTGCCTACGTGGGAT
GGCATTCGGCTCGGGGAGAGACTCCGTACCATGTCCTGTAGTGACAAAATCCTACGCTGGAACGTGCT
GGGCCTGCAAGGGGCACTGTTGACCCACTTCCTGCAGCCCATTTATCTCAAATCTGTCACATTGGGTT
ACCTTTTCAGCCAAGGGCATCTGACCCGTGCTATTTGCTGTCGTGTGACAAGAGATGGGAGTGCATTT
GAGGATGGACTACGACATCCCTTTATTGTCAACCACCCCAAGGTTGGCAGAGTCAGCATATATGATTC
CAAAAGGCAATCCGGGAAGACTAAGGAGACAAGCGTCAACTGGTGTCTGGCTGATGGCTATGACCTGG
AGATCCTGGACGGTACCAGAGGCACTGTGGATGGGCCACGGAATGAATTGTCCCGGGTCTCCAAAAAG
AACATTTTTCTTCTATTTAAGAAGCTCTGCTCCTTCCGTTACCGCAGGGATCTACTGAGACTCTCCTA
TGGTGAGGCCAAGAAAGCTGCCCGTGACTACGAGACGGCCAAGAACTACTTCAAAAAAGGCCTGAAGG
ATATGGGCTATGGGAACTGGATTAGCAAACCCCAGGAGGAAAAGAACTTTTATCTCTGCCCAGTATGA
GGGCCCATGTACGATTTAATTATGCGGACGTGATGAGCGAAGTACGATCCCACGACCGAGGCCCGTTT
AAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTG
CCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCA
TTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGG
AAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGG
GGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCG

CAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCG
CCACGTTCGCCGGTCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATA
GTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAA
ATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATA
GTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGC
AGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCT
GGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATC
GCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGG
ATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTC
CAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTAT
ATAAGCAGAGCTCTCCCTATCAGTGATAGAGATCTCCCTATCAGTGATAGAGAGCGTGCATAGGGAAC
ATCCACCACTTTAGTGAATTGTAGCACGGCTTCAGAAGCGGCCGCCCACCATGGACAAAGACTGCGAA
ATGAAGCGCACCACCCTGGATAGCCCTCTGGGCAAGCTGGAACTGTCTGGGTGCGAACAGGGCCTGCA
CCGTATCATCTTCCTGGGCAAAGGAACATCTGCCGCCGACGCCGTGGAAGTGCCTGCCCCAGCCGCCG
TGCTGGGCGGACCAGAGCCACTGATGCAGGCCACCGCCTGGCTCAACGCCTACTTTCACCAGCCTGAG
GCCATCGAGGAGTTCCCTGTGCCAGCCCTGCACCACCCAGTGTTCCAGCAGGAGAGCTTTACCCGCCA
GGTGCTGTGGAAACTGCTGAAAGTGGTGAAGTTCGGAGAGGTCATCAGCTACAGCCACCTGGCCGCCC
TGGCCGGCAATCCCGCCGCCACCGCCGCCGTGAAAACCGCCCTGAGCGGAAATCCCGTGCCCATTCTG
ATCCCCTGCCACCGGGTGGTGCAGGGCGACCTGGACGTGGGGGGCTACGAGGGCGGGCTCGCCGTGAA
AGAGTGGCTGCTGGCCCACGAGGGCCACAGACTGGGCAAGCCTGGGCTGGGTCCTGCAGGCGGAGGCG
CGCCAGGGTCTGGCGGCGGCAGTAAGAAGCTTGCCAAGGCCCGGGCTGCGCAGTCTGCCCTGGCCGCC
ATTTTTAACTTGCACTTGGATCAGACGCCATCTCGCCAGCCTATTCCCAGTGAGGGTCTTCAGCTGCA
TTTACCGCAGGTTTTAGCTGACGCTGTCTCACGCCTGGTCCTGGGTAAGTTTGGTGACCTGACCGACA
ACTTCTCCTCCCCTCACGCTCGCAGAAAAGTGCTGGCTGGAGTCGTCATGACAACAGGCACAGATGTT
AAAGATGCCAAGGTGATAAGTGTTTCTACAGGAACAAAATGTATTAATGGTGAATACATGAGTGATCG
TGGCCTTGCATTAAATGACTGCCATGCAGAAATAATATCTCGGAGATCCTTGCTCAGATTTCTTTATA
CACAACTTGAGCTTTACTTAAATAACAAAGATGATCAAAAAAGATCCATCTTTCAGAAATCAGAGCGA
GGGGGGTTTAGGCTGAAGGAGAATGTCCAGTTTCATCTGTACATCAGCACCTCTCCCTGTGGAGATGC
CAGAATCTTCTCACCACATGAGCCAATCCTGGAAGAACCAGCAGATAGACACCCAAATCGTAAAGCAA
GAGGACAGCTACGGACCAAAATAGAGTCTGGTCAGGGGACGATTCCAGTGCGCTCCAATGCGAGCATC
CAAACGTGGGACGGGGTGCTGCAAGGGGAGCGGCTGCTCACCATGTCCTGCAGTGACAAGATTGCACG
CTGGAACGTGGTGGGCATCCAGGGTTCCCTGCTCAGCATTTTCGTGGAGCCCATTTACTTCTCGAGCA
TCATCCTGGGCAGCCTTTACCACGGGGACCACCTTTCCAGGGCCATGTACCAGCGGATCTCCAACATA
GAGGACCTGCCACCTCTCTACACCCTCAACAAGCCTTTGCTCAGTGGCATCAGCAATGCAGAAGCACG
GCAGCCAGGGAAGGCCCCCAACTTCAGTGTCAACTGGACGGTAGGCGACTCCGCTATTGAGGTCATCA
ACGCCACGACTGGGAAGGATGAGCTGGGCCGCGCGTCCCGCCTGTGTAAGCACGCGTTGTACTGTCGC
TGGATGCGTGTGCACGGCAAGGTTCCCTCCCACTTACTACGCTCCAAGATTACCAAACCCAACGTGTA
CCATGAGTCCAAGCTGGCGGCAAAGGAGTACCAGGCCGCCAAGGCGCGTCTGTTCACAGCCTTCATCA
AGGCGGGGCTGGGGGCCTGGGTGGAGAAGCCCACCGAGCAGGACCAGTTCTCACTCACGCCCTGAATC
GATATTTTCAGATATCGTGTTAGTAGGGTTGCACCGACGCGCATGTGGATTAGTGCTGTGCCTTCTAG
TTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTG
TCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGT
GGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGG
CTCTATGG

Cell line **3**: CMV-enhancer – CMV promoter – $TetO_2$ – $SNAP_f$-tag – ADAR2Q – P2A – HaloTag – ADAR1Q – bGH

GACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATG
GAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATT
GACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGG
AGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATT
GACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTAC
TTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATG
GGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTG
TTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGG
CGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCCCTATCAGTGATAGAGATCTCCCTA
TCAGTGATAGAGATCGTCGACGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGAGACGCCATCCACGC
TGTTTTGACCTCCATAGAAGACACCGGGACCGATCCAGCCTCCGGACTCTAGCGTTTAAACTTAAGCT
TGGTACCGAGCTCGGATCCCCACCATGGACAAAGACTGCGAAATGAAGCGCACCACCCTGGATAGCCC
TCTGGGCAAGCTGGAACTGTCTGGGTGCGAACAGGGCCTGCACCGTATCATCTTCCTGGGCAAAGGAA
CATCTGCCGCCGACGCCGTGGAAGTGCCTGCCCCAGCCGCCGTGCTGGGCGGACCAGAGCCACTGATG
CAGGCCACCGCCTGGCTCAACGCCTACTTTCACCAGCCTGAGGCCATCGAGGAGTTCCCTGTGCCAGC
CCTGCACCACCCAGTGTTCCAGCAGGAGAGCTTTACCCGCCAGGTGCTGTGGAAACTGCTGAAAGTGG
TGAAGTTCGGAGAGGTCATCAGCTACAGCCACCTGGCCGCCCTGGCCGGCAATCCCGCCGCCACCGCC
GCCGTGAAAACCGCCCTGAGCGGAAATCCCGTGCCCATTCTGATCCCCTGCCACCGGGTGGTGCAGGG
CGACCTGGACGTGGGGGGCTACGAGGGCGGGCTCGCCGTGAAAGAGTGGCTGCTGGCCCACGAGGGCC
ACAGACTGGGCAAGCCTGGGCTGGGTCCTGCAGGCGGAGGCGCGCCAGGGTCTGGCGGCGGCAGTAAG
AAGCTTGCCAAGGCCCGGGCTGCGCAGTCTGCCCTGGCCGCCATTTTTAACTTGCACTTGGATCAGAC
GCCATCTCGCCAGCCTATTCCCAGTGAGGGTCTTCAGCTGCATTTACCGCAGGTTTTAGCTGACGCTG
TCTCACGCCTGGTCCTGGGTAAGTTTGGTGACCTGACCGACAACTTCTCCTCCCCTCACGCTCGCAGA
AAAGTGCTGGCTGGAGTCGTCATGACAACAGGCACAGATGTTAAAGATGCCAAGGTGATAAGTGTTTC
TACAGGAACAAAATGTATTAATGGTGAATACATGAGTGATCGTGGCCTTGCATTAAATGACTGCCATG
CAGAAATAATATCTCGGAGATCCTTGCTCAGATTTCTTTATACACAACTTGAGCTTTACTTAAATAAC
AAAGATGATCAAAAAAGATCCATCTTTCAGAAATCAGAGCGAGGGGGGTTTAGGCTGAAGGAGAATGT
CCAGTTTCATCTGTACATCAGCACCTCTCCCTGTGGAGATGCCAGAATCTTCTCACCACATGAGCCAA
TCCTGGAAGAACCAGCAGATAGACACCCAAATCGTAAAGCAAGAGGACAGCTACGGACCAAAATAGAG
TCTGGTCAGGGGACGATTCCAGTGCGCTCCAATGCGAGCATCCAAACGTGGGACGGGGTGCTGCAAGG
GGAGCGGCTGCTCACCATGTCCTGCAGTGACAAGATTGCACGCTGGAACGTGGTGGGCATCCAGGGTT
CCCTGCTCAGCATTTTCGTGGAGCCCATTTACTTCTCGAGCATCATCCTGGGCAGCCTTTACCACGGG
GACCACCTTTCCAGGGCCATGTACCAGCGGATCTCCAACATAGAGGACCTGCCACCTCTCTACACCCT
CAACAAGCCTTTGCTCAGTGGCATCAGCAATGCAGAAGCACGGCAGCCAGGGAAGGCCCCCAACTTCA
GTGTCAACTGGACGGTAGGCGACTCCGCTATTGAGGTCATCAACGCCACGACTGGGAAGGATGAGCTG
GGCCGCGCGTCCCGCCTGTGTAAGCACGCGTTGTACTGTCGCTGGATGCGTGTGCACGGCAAGGTTCC
CTCCCACTTACTACGCTCCAAGATTACCAAACCCAACGTGTACCATGAGTCCAAGCTGGCGGCAAAGG
AGTACCAGGCCGCCAAGGCGCGTCTGTTCACAGCCTTCATCAAGGCGGGGCTGGGGGCCTGGGTGGAG
AAGCCCACCGAGCAGGACCAGTTCTCACTCACGCCCGGCGGCCGCGGAAGCGGAGCTACTAACTTCAG
CCTGCTGAAGCAGGCTGGAGACGTGGAGGAGAACCCTGGACCTCTCGAGATGGCAGAAATCGGTACTG
GCTTTCCATTCGACCCCCATTATGTGGAAGTCCTGGGCGAGCGCATGCACTACGTCGATGTTGGTCCG
CGCGATGGCACCCCTGTGCTGTTCCTGCACGGTAACCCGACCTCCTCCTACGTGTGGCGCAACATCAT
CCCGCATGTTGCACCGACCCATCGCTGCATTGCTCCAGACCTGATCGGTATGGGCAAATCCGACAAAC
CAGACCTGGGTTATTTCTTCGACGACCACGTCCGCTTCATGGATGCCTTCATCGAAGCCCTGGGTCTG
GAAGAGGTCGTCCTGGTCATTCACGACTGGGGCTCCGCTCTGGGTTTCCACTGGGCCAAGCGCAATCC
AGAGCGCGTCAAAGGTATTGCATTTATGGAGTTCATCCGCCCTATCCCGACCTGGGACGAATGGCCAG
AATTTGCCCGCGAGACCTTCCAGGCCTTCCGCACCACCGACGTCGGCCGCAAGCTGATCATCGATCAG
AACGTTTTTATCGAGGGTACGCTGCCGATGGGTGTCGTCCGCCCGCTGACTGAAGTCGAGATGGACCA
TTACCGCGAGCCGTTCCTGAATCCTGTTGACCGCGAGCCACTGTGGCGCTTCCCAAACGAGCTGCCAA
TCGCCGGTGAGCCAGCGAACATCGTCGCGCTGGTCGAAGAATACATGGACTGGCTGCACCAGTCCCCT

GTCCCGAAGCTGCTGTTCTGGGGCACCCCAGGCGTTCTGATCCCACCGGCCGAAGCCGCTCGCCTGGC
CAAAAGCCTGCCTAACTGCAAGGCTGTGGACATCGGCCCGGGTCTGAATCTGCTGCAAGAAGACAACC
CGGACCTGATCGGCAGCGAGATCGCGCGCTGGCTGTCGACGCTGGAGATTTCCGGCCCTGCAGGCGGA
GGCGCGCCAGGGTCTGGCGGCGGCAGTAAGGCAGAACGCATGGGTTTCACAGAGGTAACCCCAGTGAC
AGGGGCCAGTCTCAGAAGAACTATGCTCCTCCTCTCAAGGTCCCCAGAAGCACAGCCAAAGACACTCC
CTCTCACTGGCAGCACCTTCCATGACCAGATAGCCATGCTGAGCCACCGGTGCTTCAACACTCTGACT
AACAGCTTCCAGCCCTCCTTGCTCGGCCGCAAGATTCTGGCCGCCATCATTATGAAAAAAGACTCTGA
GGACATGGGTGTCGTCGTCAGCTTGGGAACAGGGAATCGCTGTGTAAAAGGAGATTCTCTCAGCCTAA
AAGGAGAAACTGTCAATGACTGCCATGCAGAAATAATCTCCCGGAGAGGCTTCATCAGGTTTCTCTAC
AGTGAGTTAATGAAATACAACTCCCAGACTGCGAAGGATAGTATATTTGAACCTGCTAAGGGAGGAGA
AAAGCTCCAAATAAAAAAGACTGTGTCATTCCATCTGTATATCAGCACTGCTCCGTGTGGAGATGGCG
CCCTCTTTGACAAGTCCTGCAGCGACCGTGCTATGGAAAGCACAGAATCCCGCCACTACCCTGTCTTC
GAGAATCCCAAACAAGGAAAGCTCCGCACCAAGGTGGAGAACGGACAAGGCACAATCCCTGTGGAATC
CAGTGACATTGTGCCTACGTGGGATGGCATTCGGCTCGGGGAGAGACTCCGTACCATGTCCTGTAGTG
ACAAAATCCTACGCTGGAACGTGCTGGGCCTGCAAGGGGCACTGTTGACCCACTTCCTGCAGCCCATT
TATCTCAAATCTGTCACATTGGGTTACCTTTTCAGCCAAGGGCATCTGACCCGTGCTATTTGCTGTCG
TGTGACAAGAGATGGGAGTGCATTTGAGGATGGACTACGACATCCCTTTATTGTCAACCACCCCAAGG
TTGGCAGAGTCAGCATATATGATTCCAAAAGGCAATCCGGGAAGACTAAGGAGACAAGCGTCAACTGG
TGTCTGGCTGATGGCTATGACCTGGAGATCCTGGACGGTACCAGAGGCACTGTGGATGGGCCACGGAA
TGAATTGTCCCGGGTCTCCAAAAAGAACATTTTTCTTCTATTTAAGAAGCTCTGCTCCTTCCGTTACC
GCAGGGATCTACTGAGACTCTCCTATGGTGAGGCCAAGAAAGCTGCCCGTGACTACGAGACGGCCAAG
AACTACTTCAAAAAAGGCCTGAAGGATATGGGCTATGGGAACTGGATTAGCAAACCCCAGGAGGAAAA
GAACTTTTATCTCTGCCCAGTATGATTAATTAAGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTC
TAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCA
CTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGG
GGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGT
GGGCTCTATGG

Cell line **4**: bGH – ADAR2Q – $SNAP_f$-tag – $TetO_2$ – CMV promoter – EF1α core promoter – $TetO_2$ – HaloTag – ADAR1Q – bGH

CCATAGAGCCCACCGCATCCCCAGCATGCCTGCTATTGTCTTCCCAATCCTCCCCCTTGCTGTCCTGC
CCCACCCCACCCCCCAGAATAGAATGACACCTACTCAGACAATGCGATGCAATTTCCTCATTTTATTA
GGAAAGGACAGTGGGAGTGGCACCTTCCAGGGTCAAGGAAGGCACGGGGGAGGGGCAAACAACAGATG
GCTGGCAACTAGAAGGCACAGTCGAGGCTGATCAGCGGGTTTAAACATCGATTCAGGGCGTGAGTGAG
AACTGGTCCTGCTCGGTGGGCTTCTCCACCCAGGCCCCCAGCCCCGCCTTGATGAAGGCTGTGAACAG
ACGCGCCTTGGCGGCCTGGTACTCCTTTGCCGCCAGCTTGGACTCATGGTACACGTTGGGTTTGGTAA
TCTTGGAGCGTAGTAAGTGGGAGGGAACCTTGCCGTGCACACGCATCCAGCGACAGTACAACGCGTGC
TTACACAGGCGGGACGCGCGGCCCAGCTCATCCTTCCCAGTCGTGGCGTTGATGACCTCAATAGCGGA
GTCGCCTACCGTCCAGTTGACACTGAAGTTGGGGGCCTTCCCTGGCTGCCGTGCTTCTGCATTGCTGA
TGCCACTGAGCAAAGGCTTGTTGAGGGTGTAGAGAGGTGGCAGGTCCTCTATGTTGGAGATCCGCTGG
TACATGGCCCTGGAAAGGTGGTCCCCGTGGTAAAGGCTGCCCAGGATGATGCTCGAGAAGTAAATGGG
CTCCACGAAAATGCTGAGCAGGGAACCCTGGATGCCCACCACGTTCCAGCGTGCAATCTTGTCACTGC
AGGACATGGTGAGCAGCCGCTCCCCTTGCAGCACCCCGTCCCACGTTTGGATGCTCGCATTGGAGCGC
ACTGGAATCGTCCCCTGACCAGACTCTATTTTGGTCCGTAGCTGTCCTCTTGCTTTACGATTTGGGTG
TCTATCTGCTGGTTCTTCCAGGATTGGCTCATGTGGTGAGAAGATTCTGGCATCTCCACAGGGAGAGG
TGCTGATGTACAGATGAAACTGGACATTCTCCTTCAGCCTAAACCCCCCTCGCTCTGATTTCTGAAAG
ATGGATCTTTTTTGATCATCTTTGTTATTTAAGTAAAGCTCAAGTTGTGTATAAAGAAATCTGAGCAA
GGATCTCCGAGATATTATTTCTGCATGGCAGTCATTTAATGCAAGGCCACGATCACTCATGTATTCAC
CATTAATACATTTTGTTCCTGTAGAAACACTTATCACCTTGGCATCTTTAACATCTGTGCCTGTTGTC
ATGACGACTCCAGCCAGCACTTTTCTGCGAGCGTGAGGGGAGGAGAAGTTGTCGGTCAGGTCACCAAA

CTTACCCAGGACCAGGCGTGAGACAGCGTCAGCTAAAACCTGCGGTAAATGCAGCTGAAGACCCTCAC
TGGGAATAGGCTGGCGAGATGGCGTCTGATCCAAGTGCAAGTTAAAAATGGCGGCCAGGGCAGACTGC
GCAGCCCGGGCCTTGGCAAGCTTCTTACTGCCGCCGCCAGACCCTGGCGCGCCTCCGCCTGCAGGACC
CAGCCCAGGCTTGCCCAGTCTGTGGCCCTCGTGGGCCAGCAGCCACTCTTTCACGGCGAGCCCGCCCT
CGTAGCCCCCCACGTCCAGGTCGCCCTGCACCACCCGGTGGCAGGGGATCAGAATGGGCACGGGATTT
CCGCTCAGGGCGGTTTTCACGGCGGCGGTGGCGGCGGGATTGCCGGCCAGGGCGGCCAGGTGGCTGTA
GCTGATGACCTCTCCGAACTTCACCACTTTCAGCAGTTTCCACAGCACCTGGCGGGTAAAGCTCTCCT
GCTGGAACACTGGGTGGTGCAGGGCTGGCACAGGGAACTCCTCGATGGCCTCAGGCTGGTGAAAGTAG
GCGTTGAGCCAGGCGGTGGCCTGCATCAGTGGCTCTGGTCCGCCCAGCACGGCGGCTGGGGCAGGCAC
TTCCACGGCGTCGGCGGCAGATGTTCCTTTGCCCAGGAAGATGATACGGTGCAGGCCCTGTTCGCACC
CAGACAGTTCCAGCTTGCCCAGAGGGCTATCCAGGGTGGTGCGCTTCATTTCGCAGTCTTTGTCCATG
GTGGGCGGCCGCCCCAGAGTAAAGCTATTCGGTAATTCGTCACCCAAGAGATCAATCGGTCTCTCTCT
ATCACTGATAGGGAGATCTCTATCACTGATAGGGAGAGCTCTGCTTATATAGACCTCCCACCGTACAC
GCCTACCGCCCATTTGCGTCAATGGGGCGGAGTTGTTACGACATTTTGGAAAGTCCCGTTGATTTTGG
TGCCAAAACAAACTCCCATTGACGTCAATGGGGTGGAGACTTGGAAATCCCCGTGAGTCAAACCGCTA
TCCACGCCCATTGATGTACTGCCAAAACCGCATCACCATGGACGTGTCGAGGTGATAATTCCACTCGA
GTGGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAGGG
GTCGGCAATTGAACCGGTGCCTAGAGAAGGTGGCGCGGGGTAAACTGGGAAAGTGATGTCGTGTACTG
GCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTT
TTCGCAACGGGTTTGCCGCCAGAACACAGGTCCCTATCAGTGATAGAGATCTCCCTATCAGTGATAGA
GATCGTCGACGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGAGACGCCATCGGATCCCCACCATGGC
AGAAATCGGTACTGGCTTTCCATTCGACCCCCATTATGTGGAAGTCCTGGGCGAGCGCATGCACTACG
TCGATGTTGGTCCGCGCGATGGCACCCCTGTGCTGTTCCTGCACGGTAACCCGACCTCCTCCTACGTG
TGGCGCAACATCATCCCGCATGTTGCACCGACCCATCGCTGCATTGCTCCAGACCTGATCGGTATGGG
CAAATCCGACAAACCAGACCTGGGTTATTTCTTCGACGACCACGTCCGCTTCATGGATGCCTTCATCG
AAGCCCTGGGTCTGGAAGAGGTCGTCCTGGTCATTCACGACTGGGGCTCCGCTCTGGGTTTCCACTGG
GCCAAGCGCAATCCAGAGCGCGTCAAAGGTATTGCATTTATGGAGTTCATCCGCCCTATCCCGACCTG
GGACGAATGGCCAGAATTTGCCCGCGAGACCTTCCAGGCCTTCCGCACCACCGACGTCGGCCGCAAGC
TGATCATTGATCAGAACGTTTTTATCGAGGGTACGCTGCCGATGGGTGTCGTCCGCCCGCTGACTGAA
GTCGAGATGGACCATTACCGCGAGCCGTTCCTGAATCCTGTTGACCGCGAGCCACTGTGGCGCTTCCC
AAACGAGCTGCCAATCGCCGGTGAGCCAGCGAACATCGTCGCGCTGGTCGAAGAATACATGGACTGGC
TGCACCAGTCCCCTGTCCCGAAGCTGCTGTTCTGGGGCACCCCAGGCGTTCTGATCCCACCGGCCGAA
GCCGCTCGCCTGGCCAAAAGCCTGCCTAACTGCAAGGCTGTGGACATCGGCCCGGGTCTGAATCTGCT
GCAAGAAGACAACCCGGACCTGATCGGCAGCGAGATCGCGCGCTGGCTGTCGACGCTCGAGATTTCCG
GCCCTGCAGGCGGAGGCGCGCCAGGGTCTGGCGGCGGCAGTAAGGCAGAACGCATGGGTTTCACAGAG
GTAACCCCAGTGACAGGGGCCAGTCTCAGAAGAACTATGCTCCTCCTCTCAAGGTCCCCAGAAGCACA
GCCAAAGACACTCCCTCTCACTGGCAGCACCTTCCATGACCAGATAGCCATGCTGAGCCACCGGTGCT
TCAACACTCTGACTAACAGCTTCCAGCCCTCCTTGCTCGGCCGCAAGATTCTGGCCGCCATCATTATG
AAAAAAGACTCTGAGGACATGGGTGTCGTCGTCAGCTTGGGAACAGGGAATCGCTGTGTAAAAGGAGA
TTCTCTCAGCCTAAAAGGAGAAACTGTCAATGACTGCCATGCAGAAATAATCTCCCGGAGAGGCTTCA
TCAGGTTTCTCTACAGTGAGTTAATGAAATACAACTCCCAGACTGCGAAGGATAGTATATTTGAACCT
GCTAAGGGAGGAGAAAAGCTCCAAATAAAAAAGACTGTGTCATTCCATCTGTATATCAGCACTGCTCC
GTGTGGAGATGGCGCCCTCTTTGACAAGTCCTGCAGCGACCGTGCTATGGAAAGCACAGAATCCCGCC
ACTACCCTGTCTTCGAGAATCCCAAACAAGGAAAGCTCCGCACCAAGGTGGAGAACGGACAAGGCACA
ATCCCTGTGGAATCCAGTGACATTGTGCCTACGTGGGATGGCATTCGGCTCGGGGAGAGACTCCGTAC
CATGTCCTGTAGTGACAAAATCCTACGCTGGAACGTGCTGGGCCTGCAAGGGGCACTGTTGACCCACT
TCCTGCAGCCCATTTATCTCAAATCTGTCACATTGGGTTACCTTTTCAGCCAAGGGCATCTGACCCGT
GCTATTTGCTGTCGTGTGACAAGAGATGGGAGTGCATTTGAGGATGGACTACGACATCCCTTTATTGT
CAACCACCCCAAGGTTGGCAGAGTCAGCATATATGATTCCAAAAGGCAATCCGGGAAGACTAAGGAGA
CAAGCGTCAACTGGTGTCTGGCTGATGGCTATGACCTGGAGATCCTGGACGGTACCAGAGGCACTGTG
GATGGGCCACGGAATGAATTGTCCCGGGTCTCCAAAAAGAACATTTTTCTTCTATTTAAGAAGCTCTG
CTCCTTCCGTTACCGCAGGGATCTACTGAGACTCTCCTATGGTGAGGCCAAGAAAGCTGCCCGTGACT

ACGAGACGGCCAAGAACTACTTCAAAAAAGGCCTGAAGGATATGGGCTATGGGAACTGGATTAGCAAA
CCCCAGGAGGAAAAGAACTTTTATCTCTGCCCAGTATGATTAATTAAGTTTAAACCCGCTGATCAGCC
TCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGA
AGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTC
ATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCAT
GCTGGGGATGCGGTGGGCTCTATGG

Cell line **5**: bGH – ADAR1Q – HaloTag – $TetO_2$ – CMV promoter – EF1α core promoter – $TetO_2$ – $SNAP_f$-tag – ADAR2Q – bGH

CCATAGAGCCCACCGCATCCCCAGCATGCCTGCTATTGTCTTCCCAATCCTCCCCCTTGCTGTCCTGC
CCCACCCCACCCCCCAGAATAGAATGACACCTACTCAGACAATGCGATGCAATTTCCTCATTTTATTA
GGAAAGGACAGTGGGAGTGGCACCTTCCAGGGTCAAGGAAGGCACGGGGGAGGGGCAAACAACAGATG
GCTGGCAACTAGAAGGCACAGTCGAGGCTGATCAGCGGGTTTAAACATCGATTCATACTGGGCAGAGA
TAAAAGTTCTTTTCCTCCTGGGGTTTGCTAATCCAGTTCCCATAGCCCATATCCTTCAGGCCTTTTTT
GAAGTAGTTCTTGGCCGTCTCGTAGTCACGGGCAGCTTTCTTGGCCTCACCATAGGAGAGTCTCAGTA
GATCCCTGCGGTAACGGAAGGAGCAGAGCTTCTTAAATAGAAGAAAAATGTTCTTTTTGGAGACCCGG
GACAATTCATTCCGTGGCCCATCCACAGTGCCTCTGGTACCGTCCAGGATCTCCAGGTCATAGCCATC
AGCCAGACACCAGTTGACGCTTGTCTCCTTAGTCTTCCCGGATTGCCTTTTGGAATCATATATGCTGA
CTCTGCCAACCTTGGGGTGGTTGACAATAAAGGGATGTCGTAGTCCATCCTCAAATGCACTCCCATCT
CTTGTCACACGACAGCAAATAGCACGGGTCAGATGCCCTTGGCTGAAAAGGTAACCCAATGTGACAGA
TTTGAGATAAATGGGCTGCAGGAAGTGGGTCAACAGTGCCCCTTGCAGGCCCAGCACGTTCCAGCGTA
GGATTTTGTCACTACAGGACATGGTACGGAGTCTCTCCCCGAGCCGAATGCCATCCCACGTAGGCACA
ATGTCACTGGATTCCACAGGGATTGTGCCTTGTCCGTTCTCCACCTTGGTGCGGAGCTTTCCTTGTTT
GGGATTCTCGAAGACAGGGTAGTGGCGGGATTCTGTGCTTTCCATAGCACGGTCGCTGCAGGACTTGT
CAAAGAGGGCGCCATCTCCACACGGAGCAGTGCTGATATACAGATGGAATGACACAGTCTTTTTTATT
TGGAGCTTTTCTCCTCCCTTAGCAGGTTCAAATATACTATCCTTCGCAGTCTGGGAGTTGTATTTCAT
TAACTCACTGTAGAGAAACCTGATGAAGCCTCTCCGGGAGATTATTTCTGCATGGCAGTCATTGACAG
TTTCTCCTTTTAGGCTGAGAGAATCTCCTTTTACACAGCGATTCCCTGTTCCCAAGCTGACGACGACA
CCCATGTCCTCAGAGTCTTTTTTCATAATGATGGCGGCCAGAATCTTGCGGCCGAGCAAGGAGGGCTG
GAAGCTGTTAGTCAGAGTGTTGAAGCACCGGTGGCTCAGCATGGCTATCTGGTCATGGAAGGTGCTGC
CAGTGAGAGGGAGTGTCTTTGGCTGTGCTTCTGGGGACCTTGAGAGGAGGAGCATAGTTCTTCTGAGA
CTGGCCCCTGTCACTGGGGTTACCTCTGTGAAACCCATGCGTTCTGCCTTACTGCCGCCGCCAGACCC
TGGCGCGCCTCCGCCTGCAGGGCCGGAAATCTCGAGCGTCGACAGCCAGCGCGCGATCTCGCTGCCGA
TCAGGTCCGGGTTGTCTTCTTGCAGCAGATTCAGACCCGGGCCGATGTCCACAGCCTTGCAGTTAGGC
AGGCTTTTGGCCAGGCGAGCGGCTTCGGCCGGTGGGATCAGAACGCCTGGGGTGCCCCAGAACAGCAG
CTTCGGGACAGGGGACTGGTGCAGCCAGTCCATGTATTCTTCGACCAGCGCGACGATGTTCGCTGGCT
CACCGGCGATTGGCAGCTCGTTTGGGAAGCGCCACAGTGGCTCGCGGTCAACAGGATTCAGGAACGGC
TCGCGGTAATGGTCCATCTCGACTTCAGTCAGCGGGCGGACGACACCCATCGGCAGCGTACCCTCGAT
AAAAACGTTCTGATCAATGATCAGCTTGCGGCCGACGTCGGTGGTGCGGAAGGCCTGGAAGGTCTCGC
GGGCAAATTCTGGCCATTCGTCCCAGGTCGGGATAGGGCGGATGAACTCCATAAATGCAATACCTTTG
ACGCGCTCTGGATTGCGCTTGGCCCAGTGGAAACCCAGAGCGGAGCCCCAGTCGTGAATGACCAGGAC
GACCTCTTCCAGACCCAGGGCTTCGATGAAGGCATCCATGAAGCGGACGTGGTCGTCGAAGAAATAAC
CCAGGTCTGGTTTGTCGGATTTGCCCATACCGATCAGGTCTGGAGCAATGCAGCGATGGGTCGGTGCA
ACATGCGGGATGATGTTGCGCCACACGTAGGAGGAGGTCGGGTTACCGTGCAGGAACAGCACAGGGGT
GCCATCGCGCGGACCAACATCGACGTAGTGCATGCGCTCGCCCAGGACTTCCACATAATGGGGGTCGA
ATGGAAAGCCAGTACCGATTTCTGCCATGGTGGGCGGCCGCCCCAGAGTAAAGCTATTCGGTAATTCG
TCACCCAAGAGATCAATCGGTCTCTCTCTATCACTGATAGGGAGATCTCTATCACTGATAGGGAGAGC
TCTGCTTATATAGACCTCCCACCGTACACGCCTACCGCCCATTTGCGTCAATGGGGCGGAGTTGTTAC
GACATTTTGGAAAGTCCCGTTGATTTTGGTGCCAAAACAAACTCCCATTGACGTCAATGGGGTGGAGA
CTTGGAAATCCCCGTGAGTCAAACCGCTATCCACGCCCATTGATGTACTGCCAAAACCGCATCACCAT

GGACGTGTCGAGGTGATAATTCCACTCGAGTGGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGC
CCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGAACCGGTGCCTAGAGAAGGTGGCGCGGG
GTAAACTGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATAT
AAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGGTCCCTATCA
GTGATAGAGATCTCCCTATCAGTGATAGAGATCGTCGACGAGCTCGTTTAGTGAACCGTCAGATCGCC
TGGAGACGCCATCGGATCCCCACCATGGACAAAGACTGCGAAATGAAGCGCACCACCCTGGATAGCCC
TCTGGGCAAGCTGGAACTGTCTGGGTGCGAACAGGGCCTGCACCGTATCATCTTCCTGGGCAAAGGAA
CATCTGCCGCCGACGCCGTGGAAGTGCCTGCCCCAGCCGCCGTGCTGGGCGGACCAGAGCCACTGATG
CAGGCCACCGCCTGGCTCAACGCCTACTTTCACCAGCCTGAGGCCATCGAGGAGTTCCCTGTGCCAGC
CCTGCACCACCCAGTGTTCCAGCAGGAGAGCTTTACCCGCCAGGTGCTGTGGAAACTGCTGAAAGTGG
TGAAGTTCGGAGAGGTCATCAGCTACAGCCACCTGGCCGCCCTGGCCGGCAATCCCGCCGCCACCGCC
GCCGTGAAAACCGCCCTGAGCGGAAATCCCGTGCCCATTCTGATCCCCTGCCACCGGGTGGTGCAGGG
CGACCTGGACGTGGGGGGCTACGAGGGCGGGCTCGCCGTGAAAGAGTGGCTGCTGGCCCACGAGGGCC
ACAGACTGGGCAAGCCTGGGCTGGGTCCTGCAGGCGGAGGCGCGCCAGGGTCTGGCGGCGGCAGTAAG
AAGCTTGCCAAGGCCCGGGCTGCGCAGTCTGCCCTGGCCGCCATTTTTAACTTGCACTTGGATCAGAC
GCCATCTCGCCAGCCTATTCCCAGTGAGGGTCTTCAGCTGCATTTACCGCAGGTTTTAGCTGACGCTG
TCTCACGCCTGGTCCTGGGTAAGTTTGGTGACCTGACCGACAACTTCTCCTCCCCTCACGCTCGCAGA
AAAGTGCTGGCTGGAGTCGTCATGACAACAGGCACAGATGTTAAAGATGCCAAGGTGATAAGTGTTTC
TACAGGAACAAAATGTATTAATGGTGAATACATGAGTGATCGTGGCCTTGCATTAAATGACTGCCATG
CAGAAATAATATCTCGGAGATCCTTGCTCAGATTTCTTTATACACAACTTGAGCTTTACTTAAATAAC
AAAGATGATCAAAAAAGATCCATCTTTCAGAAATCAGAGCGAGGGGGGTTTAGGCTGAAGGAGAATGT
CCAGTTTCATCTGTACATCAGCACCTCTCCCTGTGGAGATGCCAGAATCTTCTCACCACATGAGCCAA
TCCTGGAAGAACCAGCAGATAGACACCCAAATCGTAAAGCAAGAGGACAGCTACGGACCAAAATAGAG
TCTGGTCAGGGGACGATTCCAGTGCGCTCCAATGCGAGCATCCAAACGTGGGACGGGGTGCTGCAAGG
GGAGCGGCTGCTCACCATGTCCTGCAGTGACAAGATTGCACGCTGGAACGTGGTGGGCATCCAGGGTT
CCCTGCTCAGCATTTTCGTGGAGCCCATTTACTTCTCGAGCATCATCCTGGGCAGCCTTTACCACGGG
GACCACCTTTCCAGGGCCATGTACCAGCGGATCTCCAACATAGAGGACCTGCCACCTCTCTACACCCT
CAACAAGCCTTTGCTCAGTGGCATCAGCAATGCAGAAGCACGGCAGCCAGGGAAGGCCCCCAACTTCA
GTGTCAACTGGACGGTAGGCGACTCCGCTATTGAGGTCATCAACGCCACGACTGGGAAGGATGAGCTG
GGCCGCGCGTCCCGCCTGTGTAAGCACGCGTTGTACTGTCGCTGGATGCGTGTGCACGGCAAGGTTCC
CTCCCACTTACTACGCTCCAAGATTACCAAACCCAACGTGTACCATGAGTCCAAGCTGGCGGCAAAGG
AGTACCAGGCCGCCAAGGCGCGTCTGTTCACAGCCTTCATCAAGGCGGGGCTGGGGGCCTGGGTGGAG
AAGCCCACCGAGCAGGACCAGTTCTCACTCACGCCCTGATTAATTAAGTTTAAACCCGCTGATCAGCC
TCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGA
AGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTC
ATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCAT
GCTGGGGATGCGGTGGGCTCTATGG

Constructs for HA1Q / APO1S duo cell lines 6 – 9

Cell line **6**: CMV-enhancer – CMV promoter – $TetO_2$ – mAPOBEC1 – $SNAP_f$-tag – NES – bGH – CMV-enhancer – CMV promoter – $TetO_2$ – HaloTag – ADAR1Q – bGH

GACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATG
GAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATT
GACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGG
AGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATT
GACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTAC
TTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATG
GGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTG
TTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGG
CGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCCCTATCAGTGATAGAGATCTCCCTA

TCAGTGATAGAGATCGTCGACGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGAGACGCCATCCACGC
TGTTTTGACCTCCATAGAAGACACCGGGACCGATCCAGCCTCCGGACTCTAGCGTTTAAACTTAAGCT
TGGTACCGAGCTCGGATCCGCCACCATGAGTTCCGAGACAGGCCCTGTAGCTGTTGATCCCACTCTGA
GGAGAAGAATTGAGCCCCACGAGTTTGAAGTCTTCTTTGACCCCCGGGAGCTTCGGAAAGAGACCTGT
CTGCTGTATGAGATCAACTGGGGTGGAAGGCACAGTGTCTGGCGACACACGAGCCAAAACACCAGCAA
CCACGTTGAAGTCAACTTCTTAGAAAAATTTACTACAGAAAGATACTTTCGTCCGAACACCAGATGCT
CCATTACCTGGTTCCTGTCCTGGAGTCCCTGCGGGGAGTGCTCCAGGGCCATTACAGAGTTTCTGAGC
CGACACCCCTATGTAACTCTGTTTATTTACATAGCACGGCTTTATCACCACACGGATCAGCGAAACCG
CCAAGGACTCAGGGACCTTATTAGCAGCGGTGTGACTATCCAGATCATGACAGAGCAAGAGTATTGTT
ACTGCTGGAGGAATTTCGTCAACTACCCCCCTTCAAACGAAGCATATTGGCCAAGGTACCCCCATCTG
TGGGTGAAACTGTATGTACTGGAGCTCTACTGCATCATTTTAGGACTTCCACCCTGTTTAAAAATTTT
AAGAAGAAAGCAACCTCAACTCACGTTTTTCACAATTACTCTTCAAACCTGCCATTACCAAAGGATAC
CACCCCATCTCCTTTGGGCTACAGGGTTGAAAGGAGCGGCGGCGACTGGCGCGCCAGGGCCTGCCGCG
ACTGGCGCGCCAGGGTCTGGCGGCTCCATGGACAAAGACTGCGAAATGAAGCGCACCACCCTGGATAG
CCCTCTGGGCAAGCTGGAACTGTCTGGGTGCGAACAGGGCCTGCACCGTATCATCTTCCTGGGCAAAG
GAACATCTGCCGCCGACGCCGTGGAAGTGCCTGCCCCAGCCGCCGTGCTGGGCGGACCAGAGCCACTG
ATGCAGGCCACCGCCTGGCTCAACGCCTACTTTCACCAGCCTGAGGCCATCGAGGAGTTCCCTGTGCC
AGCCCTGCACCACCCAGTGTTCCAGCAGGAGAGCTTTACCCGCCAGGTGCTGTGGAAACTGCTGAAAG
TGGTGAAGTTCGGAGAGGTCATCAGCTACAGCCACCTGGCCGCCCTGGCCGGCAATCCCGCCGCCACC
GCCGCCGTGAAAACCGCCCTGAGCGGAAATCCCGTGCCCATTCTGATCCCCTGCCACCGGGTGGTGCA
GGGCGACCTGGACGTGGGGGGCTACGAGGGCGGGCTCGCCGTGAAAGAGTGGCTGCTGGCCCACGAGG
GCCACAGACTGGGCAAGCCTGGGCTGGGTACCGGTCTGCCTCCACTTGAAAGACTGACACTGTAAGGG
CCCATGTACGATTTAATTATGCGGACGTGATGAGCGAAGTACGATCCCACGACCGAGGCCCGTTTAAA
CCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCT
TCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTG
TCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAG
ACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGC
TCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAG
CGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCA
CGTTCGCCGGTCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTA
ATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATG
GCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTA
ACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGT
ACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGC
ATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCT
ATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATT
TCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAA
AATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATA
AGCAGAGCTCTCCCTATCAGTGATAGAGATCTCCCTATCAGTGATAGAGAGCGTGCATAGGGAACATC
CACCACTTTAGTGAATTGTAGCACGGCTTCAGAAGCGGCCGCCCACCATGGCAGAAATCGGTACTGGC
TTTCCATTCGACCCCCATTATGTGGAAGTCCTGGGCGAGCGCATGCACTACGTCGATGTTGGTCCGCG
CGATGGCACCCCTGTGCTGTTCCTGCACGGTAACCCGACCTCCTCCTACGTGTGGCGCAACATCATCC
CGCATGTTGCACCGACCCATCGCTGCATTGCTCCAGACCTGATCGGTATGGGCAAATCCGACAAACCA
GACCTGGGTTATTTCTTCGACGACCACGTCCGCTTCATGGATGCCTTCATCGAAGCCCTGGGTCTGGA
AGAGGTCGTCCTGGTCATTCACGACTGGGGCTCCGCTCTGGGTTTCCACTGGGCCAAGCGCAATCCAG
AGCGCGTCAAAGGTATTGCATTTATGGAGTTCATCCGCCCTATCCCGACCTGGGACGAATGGCCAGAA
TTTGCCCGCGAGACCTTCCAGGCCTTCCGCACCACCGACGTCGGCCGCAAGCTGATCATTGATCAGAA
CGTTTTTATCGAGGGTACGCTGCCGATGGGTGTCGTCCGCCCGCTGACTGAAGTCGAGATGGACCATT
ACCGCGAGCCGTTCCTGAATCCTGTTGACCGCGAGCCACTGTGGCGCTTCCCAAACGAGCTGCCAATC
GCCGGTGAGCCAGCGAACATCGTCGCGCTGGTCGAAGAATACATGGACTGGCTGCACCAGTCCCCTGT
CCCGAAGCTGCTGTTCTGGGGCACCCCAGGCGTTCTGATCCCACCGGCCGAAGCCGCTCGCCTGGCCA
AAAGCCTGCCTAACTGCAAGGCTGTGGACATCGGCCCGGGTCTGAATCTGCTGCAAGAAGACAACCCG

GACCTGATCGGCAGCGAGATCGCGCGCTGGCTGTCGACGCTCGAGATTTCCGGCCCTGCAGGCGGAGG
CGCGCCAGGGTCTGGCGGCGGCAGTAAGGCAGAACGCATGGGTTTCACAGAGGTAACCCCAGTGACAG
GGGCCAGTCTCAGAAGAACTATGCTCCTCCTCTCAAGGTCCCCAGAAGCACAGCCAAAGACACTCCCT
CTCACTGGCAGCACCTTCCATGACCAGATAGCCATGCTGAGCCACCGGTGCTTCAACACTCTGACTAA
CAGCTTCCAGCCCTCCTTGCTCGGCCGCAAGATTCTGGCCGCCATCATTATGAAAAAAGACTCTGAGG
ACATGGGTGTCGTCGTCAGCTTGGGAACAGGGAATCGCTGTGTAAAAGGAGATTCTCTCAGCCTAAAA
GGAGAAACTGTCAATGACTGCCATGCAGAAATAATCTCCCGGAGAGGCTTCATCAGGTTTCTCTACAG
TGAGTTAATGAAATACAACTCCCAGACTGCGAAGGATAGTATATTTGAACCTGCTAAGGGAGGAGAAA
AGCTCCAAATAAAAAAGACTGTGTCATTCCATCTGTATATCAGCACTGCTCCGTGTGGAGATGGCGCC
CTCTTTGACAAGTCCTGCAGCGACCGTGCTATGGAAAGCACAGAATCCCGCCACTACCCTGTCTTCGA
GAATCCCAAACAAGGAAAGCTCCGCACCAAGGTGGAGAACGGACAAGGCACAATCCCTGTGGAATCCA
GTGACATTGTGCCTACGTGGGATGGCATTCGGCTCGGGGAGAGACTCCGTACCATGTCCTGTAGTGAC
AAAATCCTACGCTGGAACGTGCTGGGCCTGCAAGGGGCACTGTTGACCCACTTCCTGCAGCCCATTTA
TCTCAAATCTGTCACATTGGGTTACCTTTTCAGCCAAGGGCATCTGACCCGTGCTATTTGCTGTCGTG
TGACAAGAGATGGGAGTGCATTTGAGGATGGACTACGACATCCCTTTATTGTCAACCACCCCAAGGTT
GGCAGAGTCAGCATATATGATTCCAAAAGGCAATCCGGGAAGACTAAGGAGACAAGCGTCAACTGGTG
TCTGGCTGATGGCTATGACCTGGAGATCCTGGACGGTACCAGAGGCACTGTGGATGGGCCACGGAATG
AATTGTCCCGGGTCTCCAAAAAGAACATTTTTCTTCTATTTAAGAAGCTCTGCTCCTTCCGTTACCGC
AGGGATCTACTGAGACTCTCCTATGGTGAGGCCAAGAAAGCTGCCCGTGACTACGAGACGGCCAAGAA
CTACTTCAAAAAAGGCCTGAAGGATATGGGCTATGGGAACTGGATTAGCAAACCCCAGGAGGAAAAGA
ACTTTTATCTCTGCCCAGTATGAATCGATATTTTCAGATATCGTGTTAGTAGGGTTGCACCGACGCGC
ATGTGGATTAGTGCTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTT
GACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGA
GTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAAT
AGCAGGCATGCTGGGGATGCGGTGGGCTCTATGG

Cell line **7**: CMV-enhancer – CMV promoter – $TetO_2$ – HaloTag – ADAR1Q – bGH – CMV-enhancer – CMV promoter – $TetO_2$ – mAPOBEC1 – $SNAP_f$-tag – NES – bGH

GACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATG
GAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATT
GACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGG
AGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATT
GACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTAC
TTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATG
GGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTG
TTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGG
CGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCCCTATCAGTGATAGAGATCTCCCTA
TCAGTGATAGAGATCGTCGACGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGAGACGCCATCCACGC
TGTTTTGACCTCCATAGAAGACACCGGGACCGATCCAGCCTCCGGACTCTAGCGTTTAAACTTAAGCT
TGGTACCGAGCTCGGATCCCCACCATGGCAGAAATCGGTACTGGCTTTCCATTCGACCCCCATTATGT
GGAAGTCCTGGGCGAGCGCATGCACTACGTCGATGTTGGTCCGCGCGATGGCACCCCTGTGCTGTTCC
TGCACGGTAACCCGACCTCCTCCTACGTGTGGCGCAACATCATCCCGCATGTTGCACCGACCCATCGC
TGCATTGCTCCAGACCTGATCGGTATGGGCAAATCCGACAAACCAGACCTGGGTTATTTCTTCGACGA
CCACGTCCGCTTCATGGATGCCTTCATCGAAGCCCTGGGTCTGGAAGAGGTCGTCCTGGTCATTCACG
ACTGGGGCTCCGCTCTGGGTTTCCACTGGGCCAAGCGCAATCCAGAGCGCGTCAAAGGTATTGCATTT
ATGGAGTTCATCCGCCCTATCCCGACCTGGGACGAATGGCCAGAATTTGCCCGCGAGACCTTCCAGGC
CTTCCGCACCACCGACGTCGGCCGCAAGCTGATCATTGATCAGAACGTTTTTATCGAGGGTACGCTGC
CGATGGGTGTCGTCCGCCCGCTGACTGAAGTCGAGATGGACCATTACCGCGAGCCGTTCCTGAATCCT
GTTGACCGCGAGCCACTGTGGCGCTTCCCAAACGAGCTGCCAATCGCCGGTGAGCCAGCGAACATCGT
CGCGCTGGTCGAAGAATACATGGACTGGCTGCACCAGTCCCCTGTCCCGAAGCTGCTGTTCTGGGGCA

CCCCAGGCGTTCTGATCCCACCGGCCGAAGCCGCTCGCCTGGCCAAAAGCCTGCCTAACTGCAAGGCT
GTGGACATCGGCCCGGGTCTGAATCTGCTGCAAGAAGACAACCCGGACCTGATCGGCAGCGAGATCGC
GCGCTGGCTGTCGACGCTCGAGATTTCCGGCCCTGCAGGCGGAGGCGCGCCAGGGTCTGGCGGCGGCA
GTAAGGCAGAACGCATGGGTTTCACAGAGGTAACCCCAGTGACAGGGGCCAGTCTCAGAAGAACTATG
CTCCTCCTCTCAAGGTCCCCAGAAGCACAGCCAAAGACACTCCCTCTCACTGGCAGCACCTTCCATGA
CCAGATAGCCATGCTGAGCCACCGGTGCTTCAACACTCTGACTAACAGCTTCCAGCCCTCCTTGCTCG
GCCGCAAGATTCTGGCCGCCATCATTATGAAAAAAGACTCTGAGGACATGGGTGTCGTCGTCAGCTTG
GGAACAGGGAATCGCTGTGTAAAAGGAGATTCTCTCAGCCTAAAAGGAGAAACTGTCAATGACTGCCA
TGCAGAAATAATCTCCCGGAGAGGCTTCATCAGGTTTCTCTACAGTGAGTTAATGAAATACAACTCCC
AGACTGCGAAGGATAGTATATTTGAACCTGCTAAGGGAGGAGAAAAGCTCCAAATAAAAAAGACTGTG
TCATTCCATCTGTATATCAGCACTGCTCCGTGTGGAGATGGCGCCCTCTTTGACAAGTCCTGCAGCGA
CCGTGCTATGGAAAGCACAGAATCCCGCCACTACCCTGTCTTCGAGAATCCCAAACAAGGAAAGCTCC
GCACCAAGGTGGAGAACGGACAAGGCACAATCCCTGTGGAATCCAGTGACATTGTGCCTACGTGGGAT
GGCATTCGGCTCGGGGAGAGACTCCGTACCATGTCCTGTAGTGACAAAATCCTACGCTGGAACGTGCT
GGGCCTGCAAGGGGCACTGTTGACCCACTTCCTGCAGCCCATTTATCTCAAATCTGTCACATTGGGTT
ACCTTTTCAGCCAAGGGCATCTGACCCGTGCTATTTGCTGTCGTGTGACAAGAGATGGGAGTGCATTT
GAGGATGGACTACGACATCCCTTTATTGTCAACCACCCCAAGGTTGGCAGAGTCAGCATATATGATTC
CAAAAGGCAATCCGGGAAGACTAAGGAGACAAGCGTCAACTGGTGTCTGGCTGATGGCTATGACCTGG
AGATCCTGGACGGTACCAGAGGCACTGTGGATGGGCCACGGAATGAATTGTCCCGGGTCTCCAAAAAG
AACATTTTTCTTCTATTTAAGAAGCTCTGCTCCTTCCGTTACCGCAGGGATCTACTGAGACTCTCCTA
TGGTGAGGCCAAGAAAGCTGCCCGTGACTACGAGACGGCCAAGAACTACTTCAAAAAAGGCCTGAAGG
ATATGGGCTATGGGAACTGGATTAGCAAACCCCAGGAGGAAAAGAACTTTTATCTCTGCCCAGTATGA
GGGCCCATGTACGATTTAATTATGCGGACGTGATGAGCGAAGTACGATCCCACGACCGAGGCCCGTTT
AAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTG
CCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCA
TTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGG
AAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGG
GGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCG
CAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCG
CCACGTTCGCCGGTCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATA
GTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAA
ATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATA
GTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGC
AGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCT
GGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATC
GCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGG
ATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTC
CAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTAT
ATAAGCAGAGCTCTCCCTATCAGTGATAGAGATCTCCCTATCAGTGATAGAGAGCGTGCATAGGGAAC
ATCCACCACTTTAGTGAATTGTAGCACGGCTTCAGAAGCGGCCGCGCCACCATGAGTTCCGAGACAGG
CCCTGTAGCTGTTGATCCCACTCTGAGGAGAAGAATTGAGCCCCACGAGTTTGAAGTCTTCTTTGACC
CCCGGGAGCTTCGGAAAGAGACCTGTCTGCTGTATGAGATCAACTGGGGTGGAAGGCACAGTGTCTGG
CGACACACGAGCCAAAACACCAGCAACCACGTTGAAGTCAACTTCTTAGAAAAATTTACTACAGAAAG
ATACTTTCGTCCGAACACCAGATGCTCCATTACCTGGTTCCTGTCCTGGAGTCCCTGCGGGGAGTGCT
CCAGGGCCATTACAGAGTTTCTGAGCCGACACCCCTATGTAACTCTGTTTATTTACATAGCACGGCTT
TATCACCACACGGATCAGCGAAACCGCCAAGGACTCAGGGACCTTATTAGCAGCGGTGTGACTATCCA
GATCATGACAGAGCAAGAGTATTGTTACTGCTGGAGGAATTTCGTCAACTACCCCCCTTCAAACGAAG
CATATTGGCCAAGGTACCCCCATCTGTGGGTGAAACTGTATGTACTGGAGCTCTACTGCATCATTTTA
GGACTTCCACCCTGTTTAAAAATTTTAAGAAGAAAGCAACCTCAACTCACGTTTTTCACAATTACTCT
TCAAACCTGCCATTACCAAAGGATACCACCCCATCTCCTTTGGGCTACAGGGTTGAAAGGAGCGGCGG
CGACTGGCGCGCCAGGGCCTGCCGCGACTGGCGCGCCAGGGTCTGGCGGCTCCATGGACAAAGACTGC
GAAATGAAGCGCACCACCCTGGATAGCCCTCTGGGCAAGCTGGAACTGTCTGGGTGCGAACAGGGCCT

GCACCGTATCATCTTCCTGGGCAAAGGAACATCTGCCGCCGACGCCGTGGAAGTGCCTGCCCCAGCCG
CCGTGCTGGGCGGACCAGAGCCACTGATGCAGGCCACCGCCTGGCTCAACGCCTACTTTCACCAGCCT
GAGGCCATCGAGGAGTTCCCTGTGCCAGCCCTGCACCACCCAGTGTTCCAGCAGGAGAGCTTTACCCG
CCAGGTGCTGTGGAAACTGCTGAAAGTGGTGAAGTTCGGAGAGGTCATCAGCTACAGCCACCTGGCCG
CCCTGGCCGGCAATCCCGCCGCCACCGCCGCCGTGAAAACCGCCCTGAGCGGAAATCCCGTGCCCATT
CTGATCCCCTGCCACCGGGTGGTGCAGGGCGACCTGGACGTGGGGGGCTACGAGGGCGGGCTCGCCGT
GAAAGAGTGGCTGCTGGCCCACGAGGGCCACAGACTGGGCAAGCCTGGGCTGGGTACCGGTCTGCCTC
CACTTGAAAGACTGACACTGTAA

Cell line **8**: bGH – NES – $SNAP_f$-tag – mAPOBEC1 – $TetO_2$ – CMV promoter – EF1α core promoter – $TetO_2$ – HaloTag – ADAR1Q – bGH

CCATAGAGCCCACCGCATCCCCAGCATGCCTGCTATTGTCTTCCCAATCCTCCCCCTTGCTGTCCTGC
CCCACCCCACCCCCCAGAATAGAATGACACCTACTCAGACAATGCGATGCAATTTCCTCATTTTATTA
GGAAAGGACAGTGGGAGTGGCACCTTCCAGGGTCAAGGAAGGCACGGGGGAGGGGCAAACAACAGATG
GCTGGCAACTAGAAGGCACAGTCGAGGCTGATCAGCGGGTTTAAACATCGATTTACAGTGTCAGTCTT
TCAAGTGGAGGCAGACCGGTACCCAGCCCAGGCTTGCCCAGTCTGTGGCCCTCGTGGGCCAGCAGCCA
CTCTTTCACGGCGAGCCCGCCCTCGTAGCCCCCCACGTCCAGGTCGCCCTGCACCACCCGGTGGCAGG
GGATCAGAATGGGCACGGGATTTCCGCTCAGGGCGGTTTTCACGGCGGCGGTGGCGGCGGGATTGCCG
GCCAGGGCGGCCAGGTGGCTGTAGCTGATGACCTCTCCGAACTTCACCACTTTCAGCAGTTTCCACAG
CACCTGGCGGGTAAAGCTCTCCTGCTGGAACACTGGGTGGTGCAGGGCTGGCACAGGGAACTCCTCGA
TGGCCTCAGGCTGGTGAAAGTAGGCGTTGAGCCAGGCGGTGGCCTGCATCAGTGGCTCTGGTCCGCCC
AGCACGGCGGCTGGGGCAGGCACTTCCACGGCGTCGGCGGCAGATGTTCCTTTGCCCAGGAAGATGAT
ACGGTGCAGGCCCTGTTCGCACCCAGACAGTTCCAGCTTGCCCAGAGGGCTATCCAGGGTGGTGCGCT
TCATTTCGCAGTCTTTGTCCATGGAGCCGCCAGACCCTGGCGCGCCAGTCGCGGCAGGCCCTGGCGCG
CCAGTCGCCGCCGCTCCTTTCAACCCTGTAGCCCAAAGGAGATGGGGTGGTATCCTTTGGTAATGGCA
GGTTTGAAGAGTAATTGTGAAAAACGTGAGTTGAGGTTGCTTTCTTCTTAAAATTTTTAAACAGGGTG
GAAGTCCTAAAATGATGCAGTAGAGCTCCAGTACATACAGTTTCACCCACAGATGGGGGTACCTTGGC
CAATATGCTTCGTTTGAAGGGGGGTAGTTGACGAAATTCCTCCAGCAGTAACAATACTCTTGCTCTGT
CATGATCTGGATAGTCACACCGCTGCTAATAAGGTCCCTGAGTCCTTGGCGGTTTCGCTGATCCGTGT
GGTGATAAAGCCGTGCTATGTAAATAAACAGAGTTACATAGGGGTGTCGGCTCAGAAACTCTGTAATG
GCCCTGGAGCACTCCCCGCAGGGACTCCAGGACAGGAACCAGGTAATGGAGCATCTGGTGTTCGGACG
AAAGTATCTTTCTGTAGTAAATTTTTCTAAGAAGTTGACTTCAACGTGGTTGCTGGTGTTTTGGCTCG
TGTGTCGCCAGACACTGTGCCTTCCACCCCAGTTGATCTCATACAGCAGACAGGTCTCTTTCCGAAGC
TCCCGGGGGTCAAAGAAGACTTCAAACTCGTGGGGCTCAATTCTTCTCCTCAGAGTGGGATCAACAGC
TACAGGGCCTGTCTCGGAACTCATGGTGGCGCGGCCGCCCCAGAGTAAAGCTATTCGGTAATTCGTCA
CCCAAGAGATCAATCGGTCTCTCTCTATCACTGATAGGGAGATCTCTATCACTGATAGGGAGAGCTCT
GCTTATATAGACCTCCCACCGTACACGCCTACCGCCCATTTGCGTCAATGGGGCGGAGTTGTTACGAC
ATTTTGGAAAGTCCCGTTGATTTTGGTGCCAAAACAAACTCCCATTGACGTCAATGGGGTGGAGACTT
GGAAATCCCCGTGAGTCAAACCGCTATCCACGCCCATTGATGTACTGCCAAAACCGCATCACCATGGA
CGTGTCGAGGTGATAATTCCACTCGAGTGGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGCCCA
CAGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGAACCGGTGCCTAGAGAAGGTGGCGCGGGGTA
AACTGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAG
TGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGGTCCCTATCAGTG
ATAGAGATCTCCCTATCAGTGATAGAGATCGTCGACGAGCTCGTTTAGTGAACCGTCAGATCGCCTGG
AGACGCCATCGGATCCCCACCATGGCAGAAATCGGTACTGGCTTTCCATTCGACCCCCATTATGTGGA
AGTCCTGGGCGAGCGCATGCACTACGTCGATGTTGGTCCGCGCGATGGCACCCCTGTGCTGTTCCTGC
ACGGTAACCCGACCTCCTCCTACGTGTGGCGCAACATCATCCCGCATGTTGCACCGACCCATCGCTGC
ATTGCTCCAGACCTGATCGGTATGGGCAAATCCGACAAACCAGACCTGGGTTATTTCTTCGACGACCA
CGTCCGCTTCATGGATGCCTTCATCGAAGCCCTGGGTCTGGAAGAGGTCGTCCTGGTCATTCACGACT
GGGGCTCCGCTCTGGGTTTCCACTGGGCCAAGCGCAATCCAGAGCGCGTCAAAGGTATTGCATTTATG

GAGTTCATCCGCCCTATCCCGACCTGGGACGAATGGCCAGAATTTGCCCGCGAGACCTTCCAGGCCTT
CCGCACCACCGACGTCGGCCGCAAGCTGATCATTGATCAGAACGTTTTTATCGAGGGTACGCTGCCGA
TGGGTGTCGTCCGCCCGCTGACTGAAGTCGAGATGGACCATTACCGCGAGCCGTTCCTGAATCCTGTT
GACCGCGAGCCACTGTGGCGCTTCCCAAACGAGCTGCCAATCGCCGGTGAGCCAGCGAACATCGTCGC
GCTGGTCGAAGAATACATGGACTGGCTGCACCAGTCCCCTGTCCCGAAGCTGCTGTTCTGGGGCACCC
CAGGCGTTCTGATCCCACCGGCCGAAGCCGCTCGCCTGGCCAAAAGCCTGCCTAACTGCAAGGCTGTG
GACATCGGCCCGGGTCTGAATCTGCTGCAAGAAGACAACCCGGACCTGATCGGCAGCGAGATCGCGCG
CTGGCTGTCGACGCTCGAGATTTCCGGCCCTGCAGGCGGAGGCGCGCCAGGGTCTGGCGGCGGCAGTA
AGGCAGAACGCATGGGTTTCACAGAGGTAACCCCAGTGACAGGGGCCAGTCTCAGAAGAACTATGCTC
CTCCTCTCAAGGTCCCCAGAAGCACAGCCAAAGACACTCCCTCTCACTGGCAGCACCTTCCATGACCA
GATAGCCATGCTGAGCCACCGGTGCTTCAACACTCTGACTAACAGCTTCCAGCCCTCCTTGCTCGGCC
GCAAGATTCTGGCCGCCATCATTATGAAAAAAGACTCTGAGGACATGGGTGTCGTCGTCAGCTTGGGA
ACAGGGAATCGCTGTGTAAAAGGAGATTCTCTCAGCCTAAAAGGAGAAACTGTCAATGACTGCCATGC
AGAAATAATCTCCCGGAGAGGCTTCATCAGGTTTCTCTACAGTGAGTTAATGAAATACAACTCCCAGA
CTGCGAAGGATAGTATATTTGAACCTGCTAAGGGAGGAGAAAAGCTCCAAATAAAAAAGACTGTGTCA
TTCCATCTGTATATCAGCACTGCTCCGTGTGGAGATGGCGCCCTCTTTGACAAGTCCTGCAGCGACCG
TGCTATGGAAAGCACAGAATCCCGCCACTACCCTGTCTTCGAGAATCCCAAACAAGGAAAGCTCCGCA
CCAAGGTGGAGAACGGACAAGGCACAATCCCTGTGGAATCCAGTGACATTGTGCCTACGTGGGATGGC
ATTCGGCTCGGGGAGAGACTCCGTACCATGTCCTGTAGTGACAAAATCCTACGCTGGAACGTGCTGGG
CCTGCAAGGGGCACTGTTGACCCACTTCCTGCAGCCCATTTATCTCAAATCTGTCACATTGGGTTACC
TTTTCAGCCAAGGGCATCTGACCCGTGCTATTTGCTGTCGTGTGACAAGAGATGGGAGTGCATTTGAG
GATGGACTACGACATCCCTTTATTGTCAACCACCCCAAGGTTGGCAGAGTCAGCATATATGATTCCAA
AAGGCAATCCGGGAAGACTAAGGAGACAAGCGTCAACTGGTGTCTGGCTGATGGCTATGACCTGGAGA
TCCTGGACGGTACCAGAGGCACTGTGGATGGGCCACGGAATGAATTGTCCCGGGTCTCCAAAAAGAAC
ATTTTTCTTCTATTTAAGAAGCTCTGCTCCTTCCGTTACCGCAGGGATCTACTGAGACTCTCCTATGG
TGAGGCCAAGAAAGCTGCCCGTGACTACGAGACGGCCAAGAACTACTTCAAAAAAGGCCTGAAGGATA
TGGGCTATGGGAACTGGATTAGCAAACCCCAGGAGGAAAAGAACTTTTATCTCTGCCCAGTATGATTA
ATTAAGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCC
TCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAAT
TGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGG
AGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGG

Cell line **9**: bGH – ADAR1Q – HaloTag – $TetO_2$ – CMV promoter – EF1α core promoter – $TetO_2$ – mAPOBEC1 – $SNAP_f$-tag – NES – bGH

CCATAGAGCCCACCGCATCCCCAGCATGCCTGCTATTGTCTTCCCAATCCTCCCCCTTGCTGTCCTGC
CCCACCCCACCCCCCAGAATAGAATGACACCTACTCAGACAATGCGATGCAATTTCCTCATTTTATTA
GGAAAGGACAGTGGGAGTGGCACCTTCCAGGGTCAAGGAAGGCACGGGGGAGGGGCAAACAACAGATG
GCTGGCAACTAGAAGGCACAGTCGAGGCTGATCAGCGGGTTTAAACATCGATTCATACTGGGCAGAGA
TAAAAGTTCTTTTCCTCCTGGGGTTTGCTAATCCAGTTCCCATAGCCCATATCCTTCAGGCCTTTTTT
GAAGTAGTTCTTGGCCGTCTCGTAGTCACGGGCAGCTTTCTTGGCCTCACCATAGGAGAGTCTCAGTA
GATCCCTGCGGTAACGGAAGGAGCAGAGCTTCTTAAATAGAAGAAAAATGTTCTTTTTGGAGACCCGG
GACAATTCATTCCGTGGCCCATCCACAGTGCCTCTGGTACCGTCCAGGATCTCCAGGTCATAGCCATC
AGCCAGACACCAGTTGACGCTTGTCTCCTTAGTCTTCCCGGATTGCCTTTTGGAATCATATATGCTGA
CTCTGCCAACCTTGGGGTGGTTGACAATAAAGGGATGTCGTAGTCCATCCTCAAATGCACTCCCATCT
CTTGTCACACGACAGCAAATAGCACGGGTCAGATGCCCTTGGCTGAAAAGGTAACCCAATGTGACAGA
TTTGAGATAAATGGGCTGCAGGAAGTGGGTCAACAGTGCCCCTTGCAGGCCCAGCACGTTCCAGCGTA
GGATTTTGTCACTACAGGACATGGTACGGAGTCTCTCCCCGAGCCGAATGCCATCCCACGTAGGCACA
ATGTCACTGGATTCCACAGGGATTGTGCCTTGTCCGTTCTCCACCTTGGTGCGGAGCTTTCCTTGTTT
GGGATTCTCGAAGACAGGGTAGTGGCGGGATTCTGTGCTTTCCATAGCACGGTCGCTGCAGGACTTGT
CAAAGAGGGCGCCATCTCCACACGGAGCAGTGCTGATATACAGATGGAATGACACAGTCTTTTTTATT

TGGAGCTTTTCTCCTCCCTTAGCAGGTTCAAATATACTATCCTTCGCAGTCTGGGAGTTGTATTTCAT
TAACTCACTGTAGAGAAACCTGATGAAGCCTCTCCGGGAGATTATTTCTGCATGGCAGTCATTGACAG
TTTCTCCTTTTAGGCTGAGAGAATCTCCTTTTACACAGCGATTCCCTGTTCCCAAGCTGACGACGACA
CCCATGTCCTCAGAGTCTTTTTTCATAATGATGGCGGCCAGAATCTTGCGGCCGAGCAAGGAGGGCTG
GAAGCTGTTAGTCAGAGTGTTGAAGCACCGGTGGCTCAGCATGGCTATCTGGTCATGGAAGGTGCTGC
CAGTGAGAGGGAGTGTCTTTGGCTGTGCTTCTGGGGACCTTGAGAGGAGGAGCATAGTTCTTCTGAGA
CTGGCCCCTGTCACTGGGGTTACCTCTGTGAAACCCATGCGTTCTGCCTTACTGCCGCCGCCAGACCC
TGGCGCGCCTCCGCCTGCAGGGCCGGAAATCTCGAGCGTCGACAGCCAGCGCGCGATCTCGCTGCCGA
TCAGGTCCGGGTTGTCTTCTTGCAGCAGATTCAGACCCGGGCCGATGTCCACAGCCTTGCAGTTAGGC
AGGCTTTTGGCCAGGCGAGCGGCTTCGGCCGGTGGGATCAGAACGCCTGGGGTGCCCCAGAACAGCAG
CTTCGGGACAGGGGACTGGTGCAGCCAGTCCATGTATTCTTCGACCAGCGCGACGATGTTCGCTGGCT
CACCGGCGATTGGCAGCTCGTTTGGGAAGCGCCACAGTGGCTCGCGGTCAACAGGATTCAGGAACGGC
TCGCGGTAATGGTCCATCTCGACTTCAGTCAGCGGGCGGACGACACCCATCGGCAGCGTACCCTCGAT
AAAAACGTTCTGATCAATGATCAGCTTGCGGCCGACGTCGGTGGTGCGGAAGGCCTGGAAGGTCTCGC
GGGCAAATTCTGGCCATTCGTCCCAGGTCGGGATAGGGCGGATGAACTCCATAAATGCAATACCTTTG
ACGCGCTCTGGATTGCGCTTGGCCCAGTGGAAACCCAGAGCGGAGCCCCAGTCGTGAATGACCAGGAC
GACCTCTTCCAGACCCAGGGCTTCGATGAAGGCATCCATGAAGCGGACGTGGTCGTCGAAGAAATAAC
CCAGGTCTGGTTTGTCGGATTTGCCCATACCGATCAGGTCTGGAGCAATGCAGCGATGGGTCGGTGCA
ACATGCGGGATGATGTTGCGCCACACGTAGGAGGAGGTCGGGTTACCGTGCAGGAACAGCACAGGGGT
GCCATCGCGCGGACCAACATCGACGTAGTGCATGCGCTCGCCCAGGACTTCCACATAATGGGGGTCGA
ATGGAAAGCCAGTACCGATTTCTGCCATGGTGGGCGGCCGCCCCAGAGTAAAGCTATTCGGTAATTCG
TCACCCAAGAGATCAATCGGTCTCTCTCTATCACTGATAGGGAGATCTCTATCACTGATAGGGAGAGC
TCTGCTTATATAGACCTCCCACCGTACACGCCTACCGCCCATTTGCGTCAATGGGGCGGAGTTGTTAC
GACATTTTGGAAAGTCCCGTTGATTTTGGTGCCAAAACAAACTCCCATTGACGTCAATGGGGTGGAGA
CTTGGAAATCCCCGTGAGTCAAACCGCTATCCACGCCCATTGATGTACTGCCAAAACCGCATCACCAT
GGACGTGTCGAGGTGATAATTCCACTCGAGTGGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGC
CCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCAATTGAACCGGTGCCTAGAGAAGGTGGCGCGGG
GTAAACTGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATAT
AAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGGTCCCTATCA
GTGATAGAGATCTCCCTATCAGTGATAGAGATCGTCGACGAGCTCGTTTAGTGAACCGTCAGATCGCC
TGGAGACGCCATCGGATCCGCCACCATGAGTTCCGAGACAGGCCCTGTAGCTGTTGATCCCACTCTGA
GGAGAAGAATTGAGCCCCACGAGTTTGAAGTCTTCTTTGACCCCCGGGAGCTTCGGAAAGAGACCTGT
CTGCTGTATGAGATCAACTGGGGTGGAAGGCACAGTGTCTGGCGACACACGAGCCAAAACACCAGCAA
CCACGTTGAAGTCAACTTCTTAGAAAAATTTACTACAGAAAGATACTTTCGTCCGAACACCAGATGCT
CCATTACCTGGTTCCTGTCCTGGAGTCCCTGCGGGGAGTGCTCCAGGGCCATTACAGAGTTTCTGAGC
CGACACCCCTATGTAACTCTGTTTATTTACATAGCACGGCTTTATCACCACACGGATCAGCGAAACCG
CCAAGGACTCAGGGACCTTATTAGCAGCGGTGTGACTATCCAGATCATGACAGAGCAAGAGTATTGTT
ACTGCTGGAGGAATTTCGTCAACTACCCCCCTTCAAACGAAGCATATTGGCCAAGGTACCCCCATCTG
TGGGTGAAACTGTATGTACTGGAGCTCTACTGCATCATTTTAGGACTTCCACCCTGTTTAAAAATTTT
AAGAAGAAAGCAACCTCAACTCACGTTTTTCACAATTACTCTTCAAACCTGCCATTACCAAAGGATAC
CACCCCATCTCCTTTGGGCTACAGGGTTGAAAGGAGCGGCGGCGACTGGCGCGCCAGGGCCTGCCGCG
ACTGGCGCGCCAGGGTCTGGCGGCTCCATGGACAAAGACTGCGAAATGAAGCGCACCACCCTGGATAG
CCCTCTGGGCAAGCTGGAACTGTCTGGGTGCGAACAGGGCCTGCACCGTATCATCTTCCTGGGCAAAG
GAACATCTGCCGCCGACGCCGTGGAAGTGCCTGCCCCAGCCGCCGTGCTGGGCGGACCAGAGCCACTG
ATGCAGGCCACCGCCTGGCTCAACGCCTACTTTCACCAGCCTGAGGCCATCGAGGAGTTCCCTGTGCC
AGCCCTGCACCACCCAGTGTTCCAGCAGGAGAGCTTTACCCGCCAGGTGCTGTGGAAACTGCTGAAAG
TGGTGAAGTTCGGAGAGGTCATCAGCTACAGCCACCTGGCCGCCCTGGCCGGCAATCCCGCCGCCACC
GCCGCCGTGAAAACCGCCCTGAGCGGAAATCCCGTGCCCATTCTGATCCCCTGCCACCGGGTGGTGCA
GGGCGACCTGGACGTGGGGGGCTACGAGGGCGGGCTCGCCGTGAAAGAGTGGCTGCTGGCCCACGAGG
GCCACAGACTGGGCAAGCCTGGGCTGGGTACCGGTCTGCCTCCACTTGAAAGACTGACACTGTAATTA
ATTAAGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCC
TCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAAT

TGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGG
AGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGG

5.4.3. Manuscript 3 (published)

CLUSTER guideRNAs enable precise and efficient RNA editing with endogenous ADAR enzymes in vivo

Philipp Reautschnig, Nicolai Wahn, Jacqueline Wettengel, Annika E. Schulz, **Ngadhnjim Latifi**, Paul Vogel, Tae-Won Kang, Laura S. Pfeiffer, Christine Zarges, Ulrike Naumann, Lars Zender and Thorsten Stafforst, *Nature Biotechnology*, 2022, *40.5*, 759-768

nature biotechnology

ARTICLES

https://doi.org/10.1038/s41587-021-01105-0

CLUSTER guide RNAs enable precise and efficient RNA editing with endogenous ADAR enzymes in vivo

Philipp Reautschnig[1], Nicolai Wahn[2], Jacqueline Wettengel[1], Annika E. Schulz[1], Ngadhnjim Latifi[1], Paul Vogel[3], Tae-Won Kang[4,5], Laura S. Pfeiffer[1], Christine Zarges[1], Ulrike Naumann[6], Lars Zender[4,5,7], Jin Billy Li[3] and Thorsten Stafforst[1]✉

RNA base editing represents a promising alternative to genome editing. Recent approaches harness the endogenous RNA-editing enzyme adenosine deaminase acting on RNA (ADAR) to circumvent problems caused by ectopic expression of engineered editing enzymes, but suffer from sequence restriction, lack of efficiency and bystander editing. Here we present in silico-optimized CLUSTER guide RNAs that bind their target messenger RNAs in a multivalent fashion, achieve editing with high precision and efficiency and enable targeting of sequences that were not accessible using previous gRNA designs. CLUSTER gRNAs can be genetically encoded and delivered using viruses, and are active in a wide range of cell lines. In cell culture, CLUSTER gRNAs achieve on-target editing of endogenous transcripts with yields of up to 45% without bystander editing. In vivo, CLUSTER gRNAs delivered to mouse liver by hydrodynamic tail vein injection edited reporter constructs at rates of up to 10%. The CLUSTER approach opens avenues for drug development in the field of RNA base editing.

The recent development of various tools for programmable, site-specific manipulation of genetic information has created opportunities to correct disease-causing mutations. These tools include RNA-guided Cas9-nucleases and base editors that can target DNA or RNA[1]. DNA editing can induce inheritable, permanent off-target mutations, which may limit in vivo applications, whereas site-directed RNA editing alters transient RNA transcripts and may thus circumvent safety and ethics issues. Several RNA-editing tools have been engineered for targeted adenosine-to-inosine (A-to-I)[2–4] and cytidine-to-uridine (C-to-U)[5] conversion. Like DNA-editing tools, these suffer from the need to ectopically express an engineered editase[1,6], challenges in delivery and considerable off-target editing, with yet unknown consequences[4,7–10]. Indeed, ectopic expression of a highly active A-to-I editase led to lethality in a fragile murine disease model[11].

A potential solution to these concerns is to harness an endogenous editing enzyme. We[12] and others[13] have demonstrated site-directed RNA editing with the ubiquitously expressed A-to-I RNA-editing enzyme adenosine deaminase acting on RNA (ADAR)[14]. We designed a gRNA comprising a programmable antisense part for target binding (specificity domain) plus a structured RNA motif for ADAR recruitment (recruitment domain) (Fig. 1a, top)[15]. The design was particularly successful when gRNAs were administered as densely chemically modified oligonucleotides (called the RESTORE approach); editing yields in the range 20–30% have been obtained in relevant transcripts including STAT1 (ref. [12]). Notably, we found the same oligonucleotide sequence to be much less effective when expressed from a plasmid[15]. A possible explanation is the stronger binding affinity of chemically modified oligonucleotides[16]. In accordance with that, plasmid-borne gRNAs achieved successful editing of endogenous targets with endogenous ADARs when the specificity domain was extended beyond 100 nt[13]. These large, so-called LEAPER, gRNAs are fully unstructured and do not contain an ADAR recruitment domain. However, the LEAPER approach shows substantial bystander off-target editing, which is due to the formation of long, double-stranded gRNA/mRNA duplexes. Furthermore, the sequence space for the LEAPER gRNA is preassigned by the 100–150-nt sequence space at the target site[13]. Here we describe design principles for gRNAs that bind their target with a cluster of recruitment sequences (RS) freely distributed over the target. This CLUSTER design results in genetically encoded gRNAs with high sequence flexibility and enables efficient RNA editing with strongly reduced bystander editing, both in cell culture and in vivo.

Results

Design rationale and benchmarking with a previous design. The CLUSTER design is based on our previous R/G-gRNA design[15,17] but adds a cluster of single-stranded RS (Fig. 1a). In contrast to the LEAPER approach[13], which simply extends the specificity domain, the individual RS bind to the target mRNA in various regions, distal to the target site and distal to each other. The cooperative interplay of the RS together with the specificity domain was conceived to satisfy the need for a fast and strong binding of guide and target RNA while keeping the choice of the gRNA sequence highly flexible to optimize gRNA properties. To avoid bystander editing, binding sites for the RS were chosen based on the absence of highly editable adenosine bases (Methods). To facilitate the design, a custom-made

[1]Interfaculty Institute of Biochemistry, University of Tübingen, Tübingen, Germany. [2]Tübingen, Germany. [3]Department of Genetics, Stanford University, Stanford, CA, USA. [4]Department of Medical Oncology and Pneumology (Internal Medicine VIII), University Hospital Tübingen, Tübingen, Germany. [5]German Cancer Research Consortium, Partner Site Tübingen, German Cancer Research Center, Heidelberg, Germany. [6]Hertie Institute for Clinical Brain Research, Center of Neurology, University Hospital Tübingen, University of Tübingen, Tübingen, Germany. [7]DFG Cluster of Excellence 2180 'Image-guided and Functional Instructed Tumor Therapy', University of Tübingen, Tübingen, Germany. ✉e-mail: thorsten.stafforst@uni-tuebingen.de

computer program (recruitment cluster finder) was generated that searches any target RNAs for potential RS, combines them into a gRNA and finally scores them for minimal self-inhibiting secondary structure. A detailed description of the program is available in Supplementary Fig. 1.

To carve out general design rules, we edited a premature amber stop codon (5′-UAG) in a dual-luciferase reporter construct (Supplementary Fig. 2). Initially, a 20-nt-specificity domain, placing the targeted adenosine in a mismatch with cytosine (Fig. 1a), was combined with the ADAR recruiting moiety at the 5′ end and with an ensemble of three or eight RS of length 11–16 nt at the 3′ end. We tested plasmid-borne CLUSTER gRNAs in engineered 293 Flp-In T-REx cell lines stably overexpressing either ADAR1 p110, ADAR1 p150 or ADAR2, respectively (Supplementary Fig. 3 and refs. [15,17]) and compared them to previously designed gRNAs[15] with either 20- or 40-nt-specificity domains lacking RS (Fig. 1a and Supplementary Fig. 4). Both CLUSTER designs restored substantially more luciferase signal, with the 8×RS CLUSTER gRNA restoring even up to tenfold more signal in cells expressing the p110 isoform of ADAR1 (Fig. 1b). This is very desirable, as the p110 isoform is widely and highly expressed in human tissues[14,18]. In addition, ADAR2 was considerably better recruited (Supplementary Fig. 5). These trends were confirmed by Sanger sequencing at the RNA level (Fig. 1c). While LEAPER gRNAs have been described as not benefitting from an ADAR recruiting domain[13], the 5′-terminal R/G-motif fostered editing in the CLUSTER design (Fig. 1d).

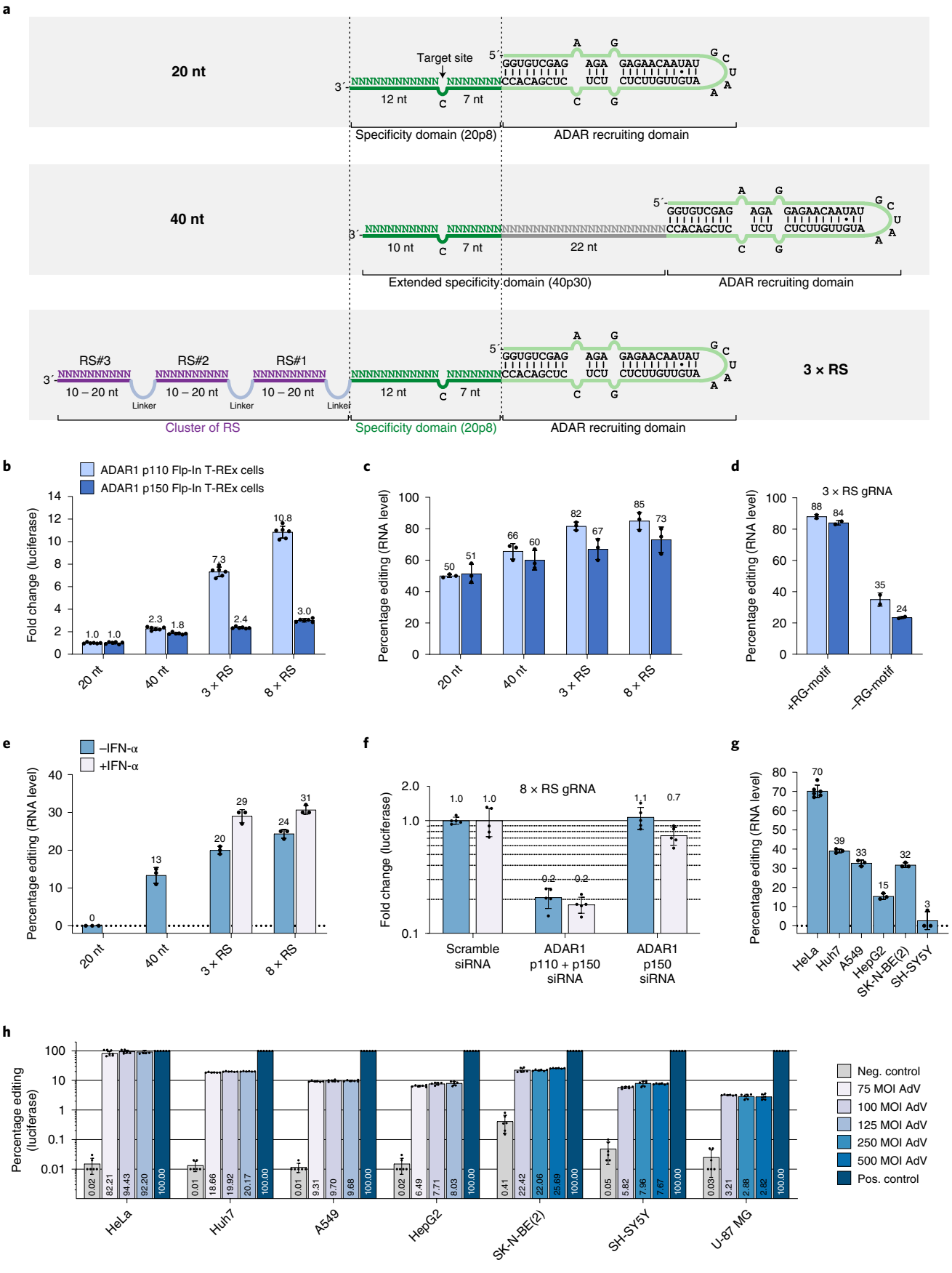
We continued the comparison of previous versus CLUSTER design for the harnessing of endogenous ADARs from HeLa cells. The prior design (20-nt-specificity domain, no RS) was unable to induce any detectable RNA editing whereas two CLUSTER gRNAs induced editing yields of 20% (3×RS) and 24% (8×RS; Fig. 1e), clearly demonstrating the strong promotional effect of RS. In contrast, simple extension of the specificity domain (from 20 to 40 nt) was less effective. To determine which endogenous ADAR induced editing, we applied small interfering RNA (siRNA) knockdown of either both ADAR1 isoforms simultaneously or of ADAR1 p150 specifically[12] (Fig. 1f). Knockdown of both isoforms gave the strongest effect (fivefold reduction of luciferase), whereas no effect was seen for the scrambled control. The effect of ADAR1 p150 knockdown on editing was comparably weak and seen only in combination with IFN-α pretreatment, which is known to induce the p150 isoform. This suggests that mainly ADAR1, predominantly isoform p110, is harnessed for editing in HeLa cells, correlating well with the expression levels of all three isoforms in this cell line[12]. Accordingly, induction of the p150 isoform with IFN-α boosted editing with CLUSTER gRNAs only slightly (Fig. 1e).

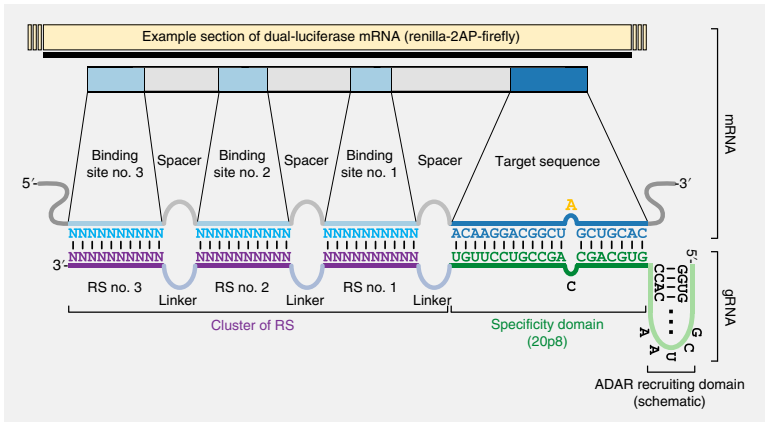
Viral delivery may overcome plasmid transfection bias in HeLa cells (Supplementary Fig. 6), and represents a promising route for the delivery of encodable gRNAs in a therapeutic context[11,19]. We generated an adenovirus expressing the 3×RS CLUSTER design from a U6 promotor, and two further adenoviruses expressing the dual-luciferase reporter either with or without a premature amber stop codon (5′-UAG), respectively. Cotransduction of reporter and 3×RS gRNA virus gave good transduction (Supplementary Fig. 6) and extensively higher editing yields (70%; Fig. 1g) compared with plasmid cotransfection of the same constructs (20%; Fig. 1e). Furthermore, luciferase signal was almost fully restored to wild-type levels (up to 94%; Fig. 1h). Compared to our previous design[16], this accounts for an efficiency increase of ~67-fold (Supplementary Fig. 7). Virus transduction enabled testing of the CLUSTER gRNA in a panel of seven difficult-to-transfect cell lines. We achieved luciferase restoration with excellent yields in HeLa (94%), good yields in Huh7, Sk-N-BE and A549 (10–20%) and moderate yields in HepG2, SH-Sy5Y and U-87 MG (3–8%). These trends were confirmed by Sanger sequencing (Fig. 1g). Although we varied the amount of gRNA virus between 75 and 500 multiplicity of infection (MOI), editing yield was hardly affected. Also, because the editing yield did not clearly correlate with expression of the reporter (Extended Data Fig. 1), of the endogenous ADAR or of the gRNA (Supplementary Fig. 6), further factors would appear to be important. Together, the data highlights that CLUSTER gRNAs recruit endogenous ADAR in various cell lines following viral delivery of gRNA.

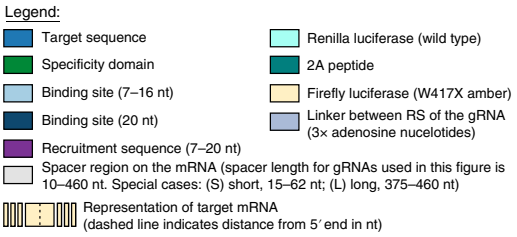
Reporter-based optimization. Next, we systematically varied the components of the CLUSTER gRNA. First, we varied the number of RS systematically between 0 and 20; at least two RS were necessary for notable editing (Fig. 2a). When increasing the number of RS from three to eight, there was only a small gain in luciferase restoration. Notably, restoration dropped slightly for the 20×RS gRNA. Thus, simply increasing the number of RS may not necessarily lead to higher editing efficiency. Unproductive folding of very large gRNAs may have been responsible for this (Extended Data Fig. 2). An alternative explanation could be the sequestration of long gRNAs by RNA-binding proteins. Based on the 3×RS CLUSTER design, we varied the size of the RS (7–15 nt) and the length of the linker separating them (0–10 nt). While the latter had only minor influence (Extended Data Fig. 3), the length of the individual RS was very important (Fig. 2e). Up to 33-fold more signal was restored with the 3×15-nt RS as compared to the 3×7-nt RS CLUSTER gRNA. Based on a 3×15-nt RS design, we systematically varied the positioning of the individual RS on the target mRNA. We found that short distances (15–62 nt) between the RS were favored over long distances (375–460 nt; Fig. 2b). However, the gRNA with the widest distribution along the target (~1,500 nt) still gave notable editing, only fivefold below the best working design in that panel. This indicates a considerable freedom of sequence selection. In another embodiment, we varied the position of the targeting sequence with respect to the RS within the gRNA. While the targeting sequence typically worked best when placed next to the ADAR recruitment domain, other placements were occasionally possible without loss of editing efficiency (Extended Data Fig. 4). Larger-sized RS (for example, >15 nt) might be helpful in boosting efficiency. To test this, we included one long RS (20 nt) in combination with two short RS (15 nt), with the RS either being in close proximity (within 170 nt) or further spread over the transcript (within 1,400 nt). We found for both distributions (proximal/distant) that the long RS should be positioned at the

Fig. 1 | CLUSTER approach compared to previous design gRNAs. a, Previous design gRNAs contained 20- or 40-nt-specificity domains combined with an ADAR recruiting domain (R/G-motif). The CLUSTER gRNA comprises additional RS of 10–20 nt in length, which bind at various regions of the target mRNA. **b**, Comparison of two previous designs (20- and 40-nt) with two CLUSTER gRNAs, each combining a 20-nt-specificity domain with either three (3×RS) or eight (8×RS) RS elements. Shown is the restoration of luciferase activity after editing of a premature 5′-UAG stop codon, reported as fold change relative to a previous gRNA design (20-nt). Experiments were carried out by plasmid cotransfection in Flp-In T-REx cells expressing specific ADAR isoforms. **c**, As in **b**, but analysis of RNA editing by RT–PCR/Sanger sequencing. **d**, The ADAR recruiting domain (R/G-motif) contributes to editing. **e**, Analysis of editing (dual-luciferase assay) in HeLa cells, harnessing endogenous ADAR. **f**, Knockdown experiment (siRNA) examining the contribution of ADAR isoforms to editing in HeLa cells, as previously established[12]. **g**, Editing of the luciferase reporter with endogenous ADARs in various cell lines with an adenovirally delivered 3×RS (15-, 13-, 11-nt) gRNA; analysis by Sanger sequencing. **h**, As in **g**, but analyzed by dual-luciferase assay (Extended Data Fig. 1). The multiplicity of infection (MOI) of the guideRNA adenovirus (AdV) is indicated in the legend. **b**–**h**, Mean ± s.d.; $n = 6$ biological replicates for **b**,**g**,**h** (HeLa cells); $n = 5$ biological replicates for **f**; $n = 3$ biological replicates for **c**,**e**,**g** (other cell lines); $n = 2$ biological replicates for **d**.

3′ end of the gRNA to maximize editing (by 1.6–2.0-fold; Fig. 2c). We tested whether the addition of stable, triplex-forming motifs[20] could enhance editing. Even though we tested six different motifs and measured editing activity 6 days after transfection, none of the triplex motifs added notable benefit (Supplementary Fig. 8). Finally, we varied the length of the specificity domain: we increased this




Example section of dual-luciferase mRNA (renilla-2AP-firefly)
mRNA
Binding site no. 3
Spacer
Binding site no. 2
Spacer
Binding site no. 1
Spacer
Target sequence
5′
3′
NNNNNNNNNN
ACAAGGACGGCU
GCUGCAC
UGUUCCUGCCGA
CGACGUG
gRNA
RS no. 3
Linker
RS no. 2
Linker
RS no. 1
Linker
Cluster of RS
Specificity domain (20p8)
ADAR recruiting domain (schematic)


Legend:
Target sequence
Specificity domain
Binding site (7–16 nt)
Binding site (20 nt)
Recruitment sequence (7–20 nt)
Spacer region on the mRNA (spacer length for gRNAs used in this figure is 10–460 nt. Special cases: (S) short, 15–62 nt; (L) long, 375–460 nt)
Representation of target mRNA (dashed line indicates distance from 5′ end in nt)
Renilla luciferase (wild type)
2A peptide
Firefly luciferase (W417X amber)
Linker between RS of the gRNA (3× adenosine nucelotides)

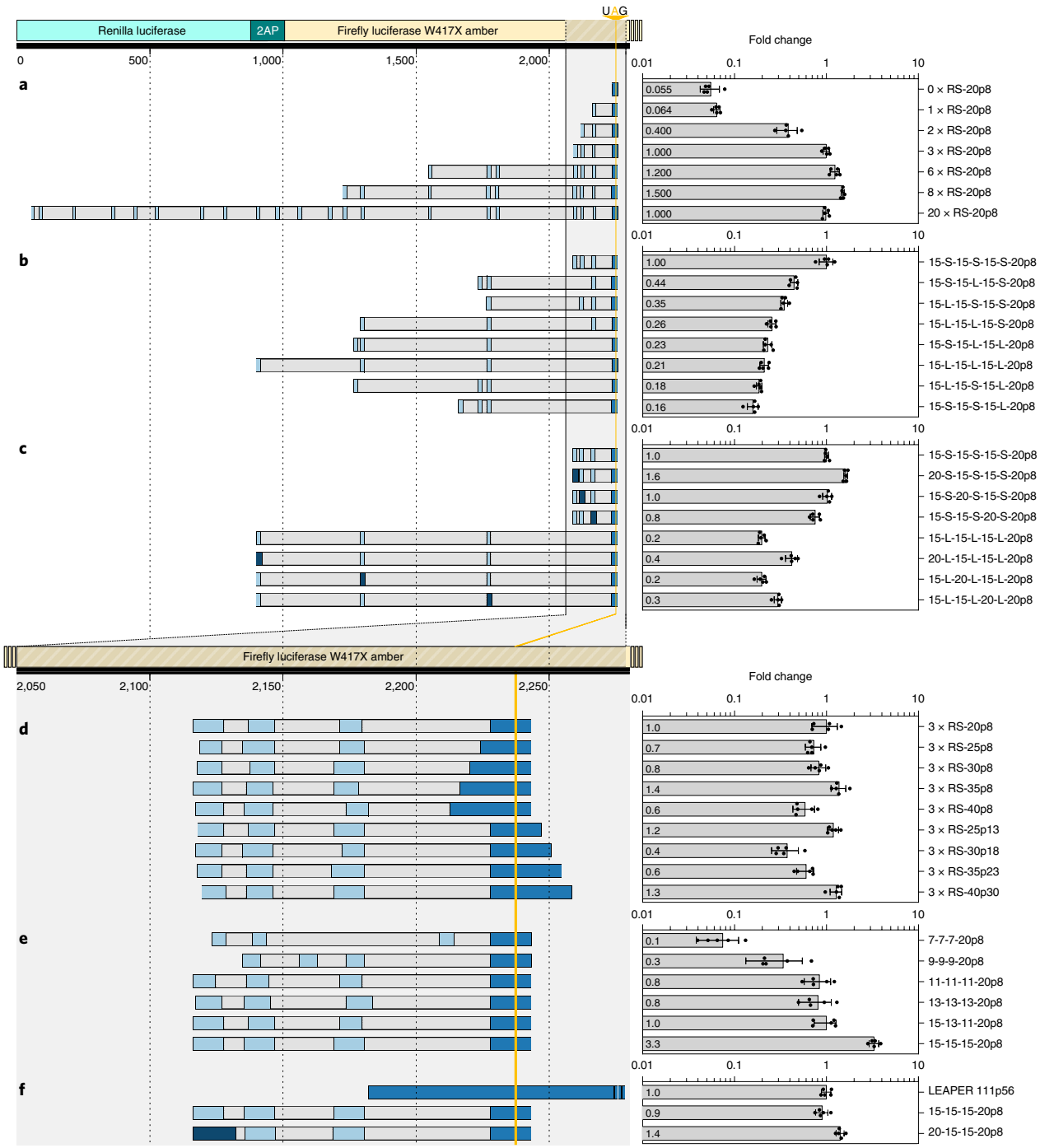

Renilla luciferase
2AP
Firefly luciferase W417X amber
UAG
Fold change
a
0 × RS-20p8
1 × RS-20p8
2 × RS-20p8
3 × RS-20p8
6 × RS-20p8
8 × RS-20p8
20 × RS-20p8
b
15-S-15-S-15-S-20p8
15-S-15-L-15-S-20p8
15-L-15-S-15-S-20p8
15-L-15-L-15-S-20p8
15-S-15-L-15-L-20p8
15-L-15-L-15-L-20p8
15-L-15-S-15-L-20p8
15-S-15-S-15-L-20p8
c
15-S-15-S-15-S-20p8
20-S-15-S-15-S-20p8
15-S-20-S-15-S-20p8
15-S-15-S-20-S-20p8
15-L-15-L-15-L-20p8
20-L-15-L-15-L-20p8
15-L-20-L-15-L-20p8
15-L-15-L-20-L-20p8
Firefly luciferase W417X amber
d
3 × RS-20p8
3 × RS-25p8
3 × RS-30p8
3 × RS-35p8
3 × RS-40p8
3 × RS-25p13
3 × RS-30p18
3 × RS-35p23
3 × RS-40p30
e
7-7-7-20p8
9-9-9-20p8
11-11-11-20p8
13-13-13-20p8
15-13-11-20p8
15-15-15-20p8
f
LEAPER 111p56
15-15-15-20p8
20-15-15-20p8

Fig. 2 | Extended design principles. a–f, CLUSTER designs, based on a 20-nt-specificity domain, were tested in HeLa cells on the dual-luciferase reporter. Editing yields are reported as fold change relative to various reference gRNAs. **a**, Influence of the number of RS elements. **b**, Influence of the positioning of RS elements. **c**, Position-specific effect of one larger RS (20 nt) in concert with two smaller RS (15 nt). The large RS is indicated by the darker color. **d**, Extending the specificity domain in 5-nt steps. **e**, Effect of the size of RS (7–15 nt) on editing efficiency. **f**, Benchmarking editing efficiency of two 3×RS CLUSTER gRNAs with LEAPER design. **a–f**, Data shown as mean ± s.d. of $n = 5$ biological replicates.

from 20 to 40 nt in 5-nt steps and varied placement of the specificity domain relative to the targeted adenosine to some extent (Fig. 2d). While extension of the specificity domain was a hallmark of the LEAPER design in improved editing[13], we could not find a clear trend for CLUSTER gRNAs. Although some of the larger-specificity domains gave slightly more luciferase signal (1.2–1.4-fold), some also gave a clear reduction (up to 2.5-fold). Both symmetric and slightly asymmetric positionings were well accepted. We recommend short specificity domains (for example, 20 nt) over longer ones to maintain bystander editing low (see below). In a first benchmark, we generated a 111-nt LEAPER gRNA against the reporter and compared it with two 3×RS CLUSTER gRNAs (Fig. 2f). The gRNA with three short RS (3×15 nt) gave editing levels comparable to the LEAPER gRNA. Notably, the CLUSTER gRNA with one larger RS (1×20 and 2×15 nt) restored roughly 1.4-fold more luciferase signal than the LEAPER gRNA, demonstrating that CLUSTER gRNAs, even though applying only a 20-nt-specificity domain, can achieve similar or even better editing yields than LEAPER gRNAs.

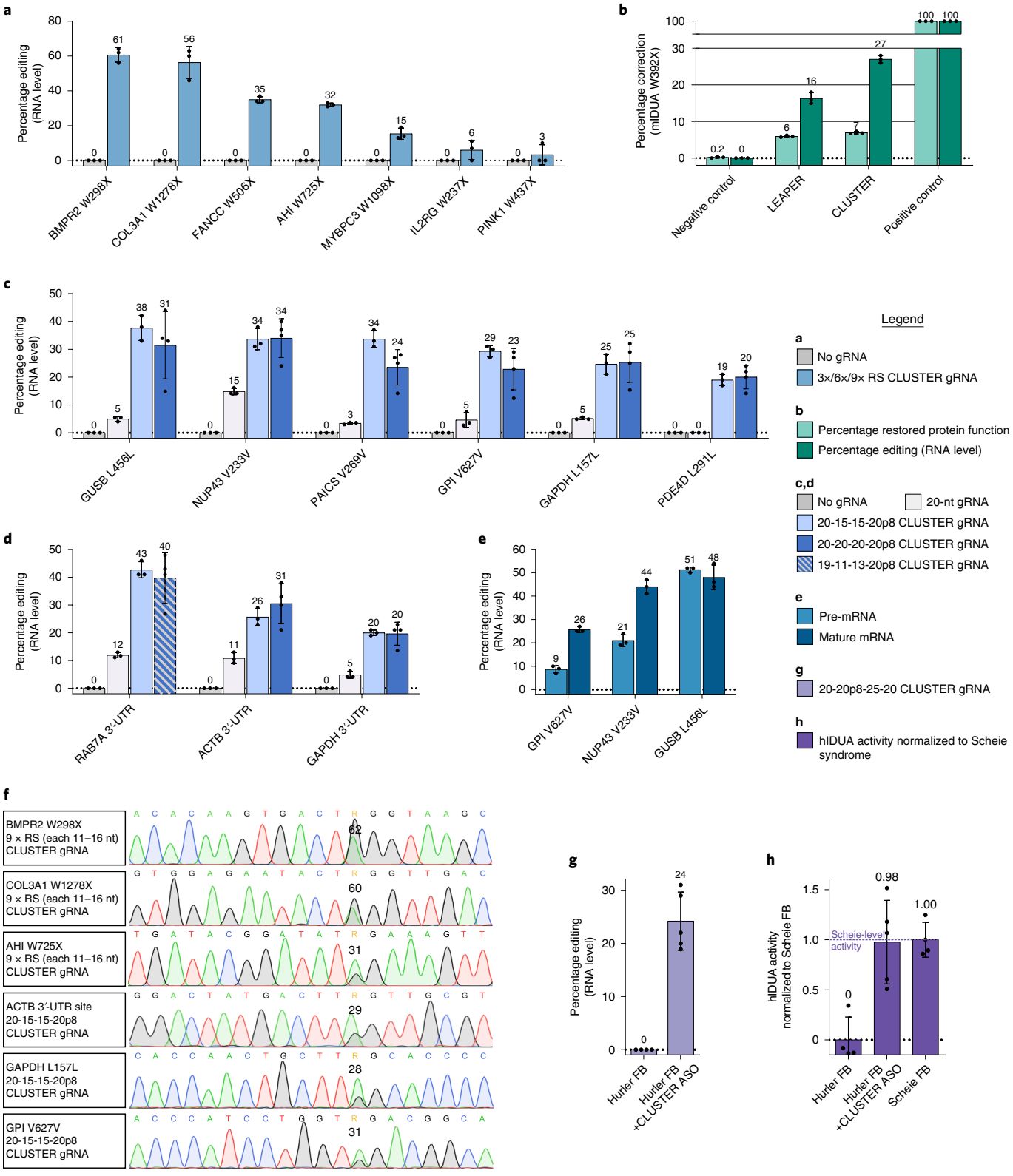
Application to disease-relevant transcripts. Premature stop codons frequently cause congenital genetic diseases. Following the derived designs rules, we constructed CLUSTER gRNAs (3–9×RS, 20-nt specificity domains) to repair nonsense mutations on eight different transcripts, each related to a specific human pathophysiology. Editing was performed in HeLa cells under plasmid cotransfection (gRNA and target complementary DNA) and was strictly gRNA dependent, resulting in editing yields between 61% and 3% (Fig. 3a,b). Notably, no bystander editing was detectable (Fig. 3f) even though adenosine-rich sequences were targeted. For two targets, BMPR2 and mIDUA, we performed a direct benchmark with the respective symmetric 111-nt LEAPER gRNA. While the LEAPER gRNA gave slightly better on-target editing than the CLUSTER gRNA on the BMPR2 target (79 versus 60%), it also showed dramatic bystander editing at 14 sites, with yields of 23–51% at ten such sites. In contrast, no bystander editing was detectable with the CLUSTER design, including regions where RS bind the target (Extended Data Fig. 5). For the mIDUA target (W392X), we found that the CLUSTER gRNA gave higher editing yields than the LEAPER gRNA (27 versus 16%). By applying a fluorogenic α-L-iduronidase assay (Fig. 3b), we benchmarked the CLUSTER and LEAPER approaches for the restoration of enzyme function: both gave similar results (7 versus 6%).

Application to endogenous transcripts. We characterized CLUSTER gRNAs on eight endogenous transcripts. These included targets with very different expression levels (for example, GAPDH: high versus GUSB, low expression) and association with human disease (for example, GUSB: mucopolysaccharidosis type VII[21], and RAB7A: Charcot–Marie–Tooth disease[22]). Specifically, we designed CLUSTER gRNAs (3×RS, 20-nt-specificity domains) targeting nine different 5′-UAG codons in eight different endogenous transcripts. We compared CLUSTER designs comprising three long RS (each 20 nt) with designs containing one long RS (3′-terminal) and two short RS (each 15 nt). Three targets were chosen in the 3′-untranslated region (UTR) and six in the open reading frame (ORF). Binding regions of all RS were localized in the exons of each respective target mRNA. Applying gRNA plasmid transfection to 293FT cells (no ADAR overexpression), we found editing yields ranging from 44 to 19% and virtually no bystander editing (Fig. 3c,d,f). Notably, the presence of the cluster of RS was always required to elicit high editing yields. Editing efficiency in ORF and 3′-UTR was similar. In the past we have repeatedly observed inefficient editing in the ORF compared to 3′-UTR, and found that global blockage of translation boosts editing in the ORF[8,23]. However, the combined data here suggest that editing with CLUSTER gRNAs takes place in the nucleus, where U6 promotor-driven, plasmid-borne gRNAs reside (Supplementary Fig. 9). In addition, absence of the effect of IFN-α treatment and of ADAR1 p150-specific knockdown on editing fits to a model where editing takes place in the nucleus, depending mostly on nuclear ADAR1 p110. Thus, we tested editing yields in the pre-mRNA of three targets (GPI, NUP43, GUSB) and compared these to editing levels in mature mRNAs. For GUSB, the editing levels before and after splicing were equal, which supports a model where editing occurs entirely in the nucleus (Fig. 3e). In the two other targets, the editing levels in the pre-mRNA were one-third to one-half of the final editing yield, supporting a model in which editing must have taken place in the nucleus at least partly.

Finally, we tested the CLUSTER approach for restoration of hIDUA activity in fibroblasts taken from a patient with Hurler syndrome. To overcome the strong plasmid transfection bias in these cells, we applied the CLUSTER gRNA in the form of an antisense oligonucleotide, very similar to a procedure previously done with LEAPER gRNAs[13]. Using Sanger sequencing, we determined gRNA-dependent editing yields of 24% (Fig. 3g) and observed restoration of IDUA enzyme activity to the levels obtained with Scheie control fibroblasts (Fig. 3h). Because Scheie syndrome is much less severe, the data indicate that clinical benefit may be within reach. Overall, RNA and enzyme assay data match those of previous LEAPER gRNAs[13], indicating again that CLUSTER and LEAPER designs perform similarly regarding efficiency.

Fig. 3 | Application to endogenous and disease-relevant targets. a, CLUSTER gRNAs with 3×RS (PINK1), 6×RS (IL2RG) or 9×RS (other) were applied to correct nonsense mutations on various disease-relevant targets (cDNAs) in HeLa cells. **b**, Benchmark of a 3×RS CLUSTER and a 111-nt LEAPER design for the correction of a disease-relevant nonsense mutation in mIDUA. Protein restoration was analyzed using an α-L-iduronidase assay. **c**, Editing of 5′-UAG codons in the ORF of various endogenous transcripts by harnessing of endogenous ADAR in 293FT cells. Targets cover both highly and weakly expressed transcripts, as well as disease-relevant transcripts (GUSB, RAB7A). Controls served either no gRNAs or gRNAs of previous design lacking RS elements. **d**, As in **c**, but editing of 5′-UAG codons in the 3′-UTR of different endogenous transcripts. **e**, Assessment of editing yields in spliced and unspliced mRNAs when targeting the ORF of three endogenous transcripts with endogenous ADAR in 293FT cells. **f**, Example Sanger sequencing reads from **a**,**c**,**d** demonstrate the lack of bystander editing. **g**, Correction of the disease-relevant hIDUA W402amber mutation in primary fibroblasts (FB) from a patient with Hurler syndrome (Hurler FB, GM06214) via transfection of a chemically synthesized CLUSTER antisense oligonucleotide (ASO). Restoration of protein function was determined using the α-L-iduronidase assay. IDUA enzyme activity in Hurler fibroblasts was normalized to IDUA activity measured in fibroblasts from a patient with Scheie syndrome (Scheie FB, GM01323). **a**–**e**,**g**, Data are mean ± s.d., with $n = 3$ biological replicates (**a**,**b**,**e**), $n \geq 3$ biological replicates (**c**,**d**) and $n \geq 4$ biological replicates (**g**).

Benchmarking of CLUSTER and LEAPER approaches. For benchmarking we designed symmetric 111-nt LEAPER gRNAs against two endogenous ORF targets (GUSB and NUP43) and one 3′-UTR target (RAB7A), and tested them side by side with CLUSTER designs. Although the LEAPER gRNA for NUP43 V233V achieved similar editing yields (46%) compared with the CLUSTER design (43%), with the former we detected seven bystander off-target events with yields ranging from 10 to 33% (Fig. 4a). In contrast, the best CLUSTER gRNA elicited only minimal off-target editing (4%) at one site. In accordance with our computational design, there was no bystander editing at or around the binding sites of all three RS. Others have suggested that suppression of bystander editing caused by LEAPER gRNAs can be achieved by putting off-target adenosines into mismatch with guanosines[13]. Thus, we generated a 111-nt LEAPER gRNA with six A:G mismatches at the worst bystander sites. However, this was not successful. Bystander editing was still



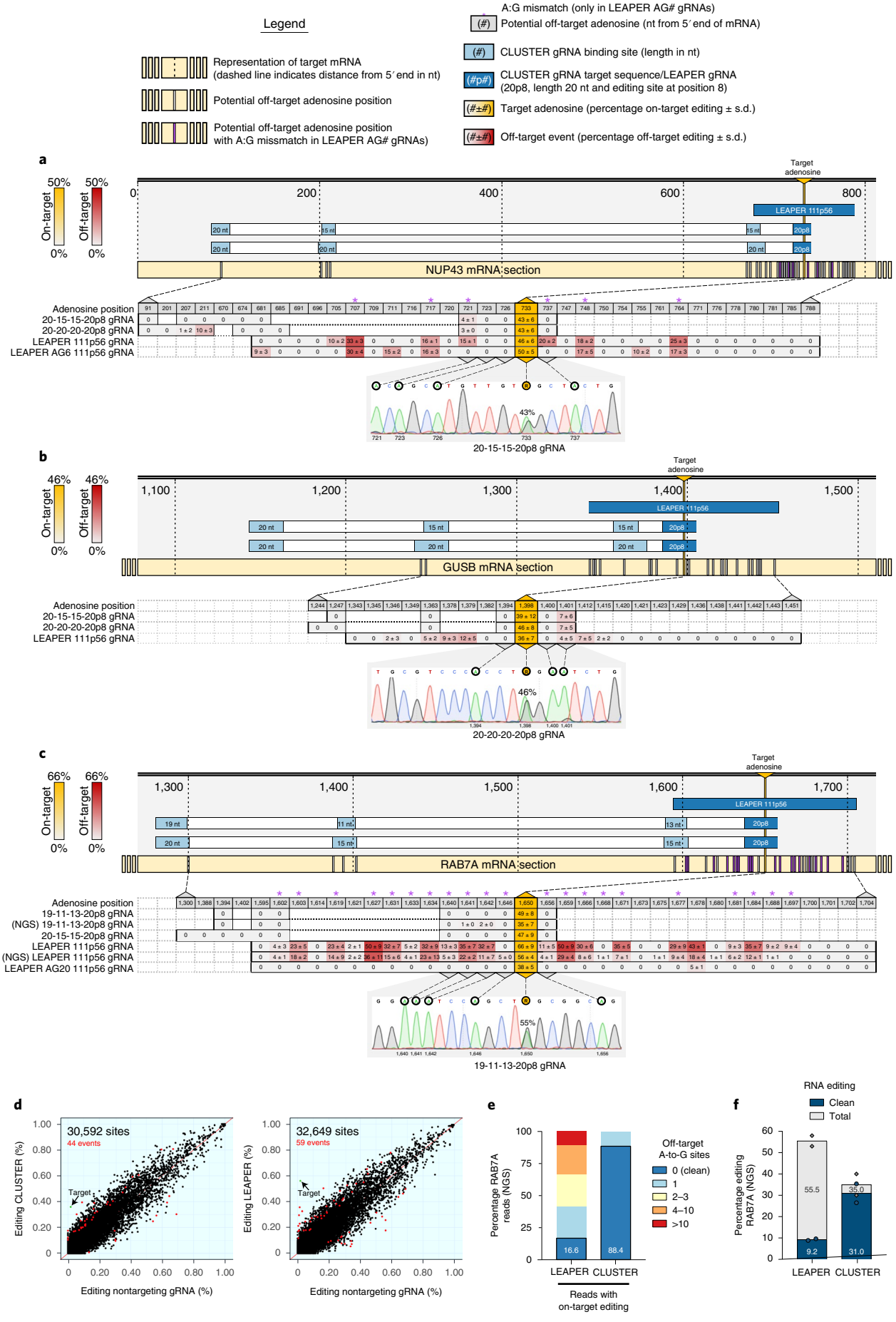

Legend
Representation of target mRNA (dashed line indicates distance from 5′ end in nt)
Potential off-target adenosine position
Potential off-target adenosine position with A:G missmatch in LEAPER AG# gRNAs)
A:G mismatch (only in LEAPER AG# gRNAs)
(#) Potential off-target adenosine (nt from 5′ end of mRNA)
(#) CLUSTER gRNA binding site (length in nt)
(#p#) CLUSTER gRNA target sequence/LEAPER gRNA (20p8, length 20 nt and editing site at position 8)
(#±#) Target adenosine (percentage on-target editing ± s.d.)
(#±#) Off-target event (percentage off-target editing ± s.d.)
a
On-target
Off-target
Target adenosine
LEAPER 111p56
NUP43 mRNA section
Adenosine position
20-15-15-20p8 gRNA
20-20-20-20p8 gRNA
LEAPER 111p56 gRNA
LEAPER AG6 111p56 gRNA
43%
20-15-15-20p8 gRNA
b
GUSB mRNA section
46%
20-20-20-20p8 gRNA
c
RAB7A mRNA section
19-11-13-20p8 gRNA
(NGS) 19-11-13-20p8 gRNA
(NGS) LEAPER 111p56 gRNA
LEAPER AG20 111p56 gRNA
55%
19-11-13-20p8 gRNA
d
30,592 sites
44 events
32,649 sites
59 events
Target
Editing CLUSTER (%)
Editing LEAPER (%)
Editing nontargeting gRNA (%)
e
Percentage RAB7A reads (NGS)
Off-target A-to-G sites
0 (clean)
1
2–3
4–10
>10
16.6
88.4
LEAPER CLUSTER
Reads with on-target editing
f
RNA editing
Clean
Total
Percentage editing RAB7A (NGS)
55.5
9.2
35.0
31.0
LEAPER CLUSTER

Fig. 4 | Benchmarking of CLUSTER and LEAPER designs. a, Two CLUSTER gRNAs (3×RS, 20-nt-specificity domain) compared with two 111-nt LEAPER gRNAs containing either no or six A:G mismatches to control bystander editing. Editing was performed with plasmid-borne gRNA in 293FT cells (endogenous ADAR) to edit a 5′-UAG codon in the endogenous NUP43 transcript. On-target and potential bystander sites are indicated. A representative Sanger trace for a CLUSTER gRNA is shown. **b**, As in **a** but with three gRNAs targeting the ORF of GUSB. **c**, As in **a** but with four gRNAs targeting the 3′-UTR of RAB7A. Editing yields for one CLUSTER and one LEAPER gRNA were additionally analyzed by NGS. **d**, Analysis of global off-target editing (poly(A) + transcriptome) when recruiting endogenous ADAR to edit a 5′-UAG site in endogenous RAB7A either with a CLUSTER (3×RS) or LEAPER (111-nt) gRNA. Scatter plots show differential editing at ~30,000 sites comparing editing levels in cells transfected with plasmids carrying either the CLUSTER, the LEAPER or a nontargeting gRNA, respectively. Experiments were done with two independent replicates. Arrow indicates on-target editing. Significantly differently edited sites (adjusted $P < 0.01$, Fisher's exact test, two-tailed, $n \geq 50$) are highlighted in red. **e**, Assessment of editing precision: all NGS reads with on-target editing were checked for bystander editing. While CLUSTER gRNAs gave mainly clean sequencing reads, LEAPER samples very frequently contained several bystander edits. **f**, Estimation of clean editing yield (no bystander) versus total on-target editing. **a**–**c**, On-target and bystander editing yields shown as mean ± s.d. of $n = 3$ biological replicates. **c**–**f**, NGS analyses and editing yields based on results from $n = 2$ biological replicates.

detectable and additional bystander sites appeared while others disappeared. Although the LEAPER gRNA against the 3′-UTR of RAB7A gave excellent editing yields, it was impaired by major off-target editing at >20 bystander sites, with yields of up to 50% at several sites (Fig. 4c). Again, we generated a LEAPER variant, this time with 20 A:G mismatches. While bystander editing could be solved in this case, on-target yield dropped dramatically (down to 38%). In contrast, both CLUSTER gRNAs gave excellent on-target yields (47–49%) with no detectable bystander editing. Finally, we generated a LEAPER gRNA against GUSB L456L that was moderately effective, giving 36% editing yield and moderate bystander editing at five sites with yields ranging from 4 to 12%. In contrast, both CLUSTER gRNAs gave better editing yields and less bystander editing. The CLUSTER design with long RS (3×20 nt) gave the best result, with 46% on-target and one bystander editing (7%). Together, these data show that CLUSTER gRNAs can achieve editing levels similar to, or even better than, LEAPER gRNAs but offer reliable control of bystander editing.

To compare global off-target effects and low-level bystander editing between the LEAPER and CLUSTER approaches, we performed a transcriptome-wide poly(A) + RNA-sequencing (RNA-seq) experiment for the RAB7A target in HEK293FT cells. A nontargeting gRNA was transfected as control. We used our established pipeline[12], to identify significantly differently edited sites in the transcriptome, and found a small number of hits for the LEAPER (59) and the CLUSTER (44) gRNA (Fig. 4d), respectively. However, a large fraction of these hits were known editing sites in Alu repeats. The number of novel sites that had not been edited in the nontargeting control was only three for the CLUSTER gRNA (apart from the target site) and seven for the LEAPER (Extended Data Fig. 6). For the CLUSTER gRNA, two of the three novel sites were exonic (HTATSF1 and CTNNAL1), both containing putative binding sites for the gRNA and giving low editing yields (≤12.2%). The HTATSF1 editing site was the only missense mutation. Notably, the LEAPER gRNA edited that site to a similar extent. The other seven novel sites detected with the LEAPER gRNA were bystander editing in close proximity to the on-target site. To analyze bystander editing more comprehensively, we looked specifically into the binding regions of the LEAPER and CLUSTER gRNA at the RAB7A transcript. In accordance with Sanger sequencing, we identified bystander editing at >20 sites for the LEAPER gRNA, ten with yields of 12–36%. In contrast, the CLUSTER gRNA gave almost no bystander editing, with only two low-level events around the target site (1–2%). In particular, the binding regions of RS showed no detectable bystander editing, highlighting the functioning of the computational design process. Next-generation sequencing (NGS) allows study of the frequency by which on-target editing is corrupted with bystander editing within the same read. While 88.4% of reads with the CLUSTER design were cleanly edited, this was the case for only 16.6% in the LEAPER gRNA (Fig. 4e and Extended Data Fig. 7). If one factors this in, the total yield of clean on-target editing was 31% for CLUSTER gRNA, clearly outcompeting LEAPER, which achieved only 9.2% (Fig. 4f). Furthermore, the NGS data revealed that neither the expression level of the target transcript (Extended Data Fig. 8) nor exon usage (Extended Data Fig. 9) was significantly affected during editing.

In vivo application. Finally, we aimed to demonstrate the recruitment of endogenous ADAR in vivo. For this, we codelivered the dual-luciferase reporter plasmid together with the plasmid-borne 3×RS CLUSTER gRNA from our first round (3×RS: 15-, 13- and 11 + 20-nt-specificity domains) into wild-type C57BL/6 mice by hydrodynamic tail vein injection[24] (Fig. 5a and Supplementary Fig. 10). Mice were euthanized and livers harvested at 72 h post injection, and RNA editing was analyzed by luminescence detection and Sanger sequencing. We found strictly gRNA-dependent editing (Fig. 5a). Both assays gave similar editing yields of around 5%. Next, we tested the gRNA design that resulted from our optimization round (3×RS: 20-, 15- and 15 + 20-nt-specificity domains). In HeLa cells, the latter design outperformed the earlier design 1.8-fold (Fig. 5b). In accordance with that, the 20-15-15 design induced twofold better editing in BL6 mice yielding 10% restoration, as quantified by Sanger sequencing and luminescence detection (Fig. 5a). Notably, no bystander editing was observed with either type of gRNA. This demonstrates significant recruitment of endogenous ADAR with a genetically encoded gRNA in vivo.

Discussion

Our previous gRNA design[15,17] has recently been applied by others[11] in two murine disease models in vivo. Consistent with our experience in cell culture (Fig. 5b)[15], coexpression of ADAR enzymes was required to achieve editing yields >0.6% (ref. [11]). However, overexpression of ADAR enzymes—in particular the hyperactive ADAR2 mutant E488Q—led to massive global off-target editing and lethality[11]. Here we demonstrate that the addition of a cluster of recruitment sequences increases editing efficiency ~30-fold (Fig. 5b) and enables editing of a reporter in vivo with yields of up to 10% without the need to coexpress an editing enzyme (Fig. 5a). This increase in editing efficiency is not explained by upregulation of ADAR expression, which was observed neither in vivo (Supplementary Fig. 11) nor in HeLa (Supplementary Fig. 12) or HEK293FT cells (Extended Data Fig. 8). CLUSTER gRNAs differ from competing approaches. The RESTORE approach applies chemical modifications[16] to circumvent bystander editing and to increase on-target efficiency[12]. However, such heavily chemically modified antisense oligonucleotides (ASO) cannot be delivered in a genetically encoded manner, such as in viruses, which may limit their effect duration[16]. Furthermore, RESTORE ASOs currently give comparably moderate editing yields and require the expression of the ADAR1 p150 isoform—for example, via IFN-α induction[12]. On the other hand, the LEAPER approach achieves the harnessing of endogenous ADARs with genetically encoded gRNAs[13]; however, the sequence space is

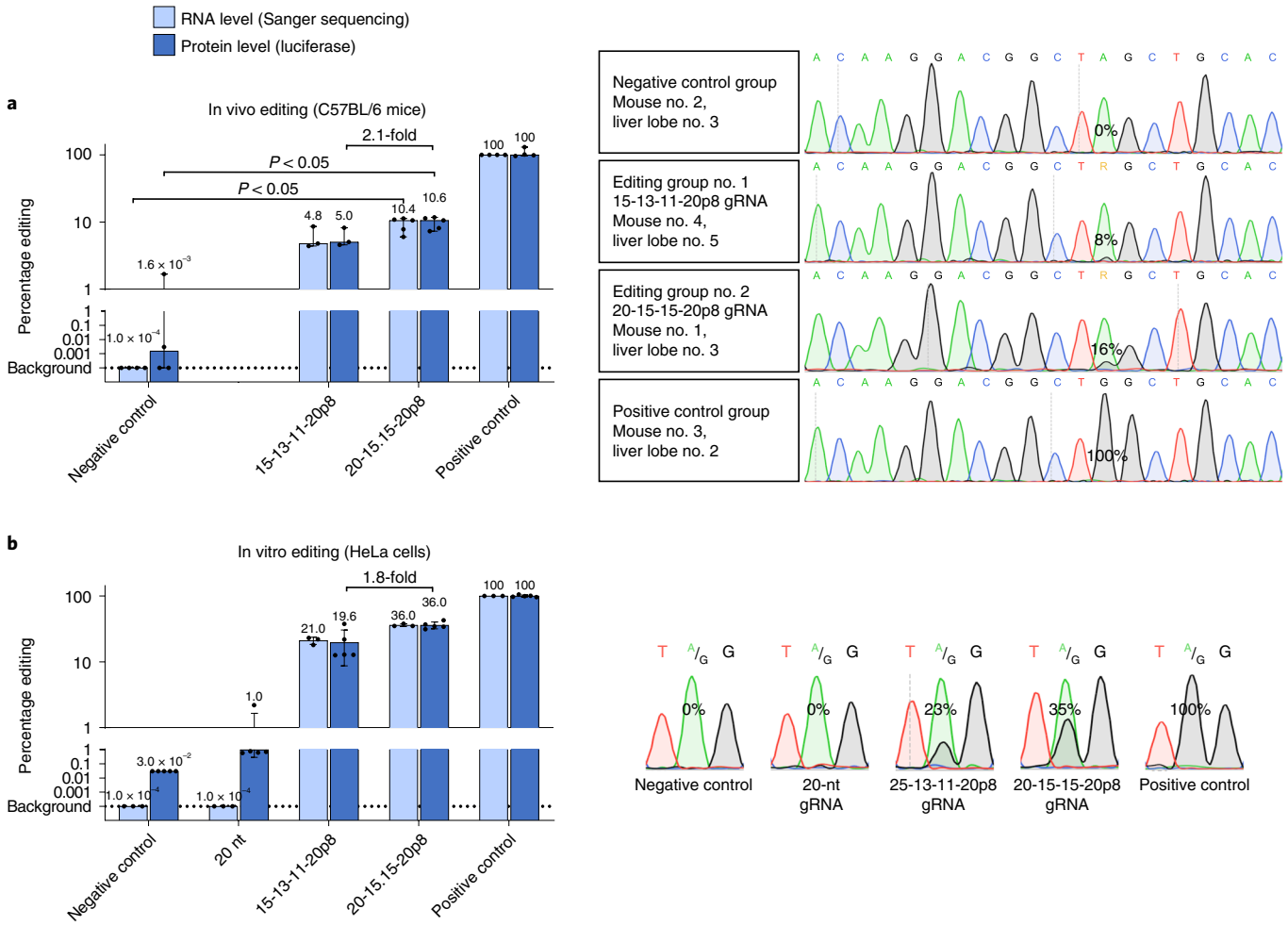


Fig. 5 | In vivo application. a, Two CLUSTER gRNAs (3 × RS of different length, 20-nt-specificity domain) were tested for in vivo restoration of luciferase activity in C57BL/6 mice. Dual-luciferase reporter and gRNA plasmids were codelivered by hydrodynamic tail vein injection. Luciferase signal was normalized to that in the positive control. As a negative control, no gRNA was applied. Representative Sanger traces are shown. **b**, In vitro experiment with the same CLUSTER constructs in HeLa cells, including the previous design of gRNA with short specificity domain (20-nt) lacking RS elements. For statistical analysis, a Mann–Whitney *U*-test (two-tailed, nonparametric) was applied. **a**, Editing yields are median (confidence interval 95%) determined in $n = 3$ mice (15-13-11-20p8 group), $n = 4$ mice (control groups) and $n = 5$ mice (20-15-15-20p8 group) measured in technical replicates as described in Methods. **b**, Editing yields shown are mean ± s.d. of $n = 3$ (Sanger) or $n = 5$ (luciferase) biological replicates.

strongly restricted and the length of the gRNAs (70–150 nt) causes problems. The formation of long, double-stranded RNA structures leads to massive bystander off-target editing[13], which cannot always be easily controlled (Fig. 4). Furthermore, the long LEAPER gRNAs could lead to immune induction and interference with splicing and translation.

In contrast, the CLUSTER design opens a large sequence space for optimization. While we have provided some general design rules, finding the optimal CLUSTER gRNA for any given target might require some screening, which can be assisted by current and future in silico methods. In our current algorithm, the avoidance of bystander off-target editing and nonproductive folding was included. Accordingly, CLUSTER gRNAs achieved good editing yields with very little bystander editing, even when the RS were distributed over 1,000 and more nucleotides, demonstrating their advantage over the less flexible LEAPER design. Notably, the algorithm could be further extended in the future to avoid immune-stimulating oligonucleotide sequences[25]—for example, motifs that have been described to induce Toll-like or RIG-I-like receptor-mediated signaling—or to remove potential transcriptome-wide off-target editing due to misguiding by the gRNA. However, NGS analysis of the specific example above (Fig. 4d) revealed no strong necessity for the latter. Finally, the concept of CLUSTER gRNAs could be extended to other platforms for RNA editing and modification, including artificial A-to-I and C-to-U editing approaches based on λN-ADAR[3], MCP-ADAR[11] or dCas13-ADAR[4,5]. However, with respect to clinical application, the harnessing of endogenous ADARs with genetically encoded, virally delivered CLUSTER gRNAs appears most promising[26]. Nevertheless, administration of gRNAs in a chemically modified format may also be successful, as shown here in Hurler syndrome fibroblasts[12].

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41587-021-01105-0.

Received: 27 October 2020; Accepted: 23 September 2021;
Published online: 03 January 2022

References

1. Rees, H. A. & Liu, D. R. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.* **19**, 770–788 (2018).
2. Stafforst, T. & Schneider, M. F. An RNA-deaminase conjugate selectively repairs point mutations. *Angew. Chem. Int. Ed. Engl.* **51**, 11166–11169 (2012).
3. Montiel-Gonzalez, M. F. et al. Correction of mutations within the cystic fibrosis transmembrane conductance regulator by site-directed RNA editing. *Proc. Natl Acad. Sci. USA* **110**, 18285–18290 (2013).
4. Cox, D. B. T. et al. RNA editing with CRISPR-Cas13. *Science* **358**, 1019–1027 (2017).
5. Abudayyeh, O. O. et al. A cytosine deaminase for programmable single-base RNA editing. *Science* **365**, 382–386 (2019).
6. Vogel, P. & Stafforst, T. Critical review on engineering deaminases for site-directed RNA editing. *Curr. Opin. Biotechnol.* **55**, 74–80 (2018).
7. Vallecillo-Viejo, I. C. et al., Abundant off-target edits from site-directed RNA editing can be reduced by nuclear localization of the editing enzyme. *RNA Biol.* **15**, 104–114 (2017).
8. Vogel, P. et al. Efficient and precise editing of endogenous transcripts with SNAP-tagged ADARs. *Nat. Methods* **15**, 535–538 (2018).
9. Grunewald, J. et al. Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature* **569**, 433–437 (2019).
10. Kim, D. et al. Genome-wide target specificity of CRISPR RNA-guided adenine base editors. *Nat. Biotechnol.* **37**, 430–435 (2019).
11. Katrekar, D. et al.. In vivo RNA editing of point mutations via RNA-guided adenosine deaminases. *Nat. Methods* **16**, 239–242 (2019).
12. Merkle, T. et al. Precise RNA editing by recruiting endogenous ADARs with antisense oligonucleotides. *Nat. Biotechnol.* **37**, 133–138 (2019).
13. Qu, L. et al. Programmable RNA editing by recruiting endogenous ADAR using engineered RNAs. *Nat. Biotechnol.* **37**, 1059–1069 (2019).
14. Nishikura, K. Functions and regulation of RNA editing by ADAR deaminases. *Annu. Rev. Biochem.* **79**, 321–349 (2010).
15. Wettengel, J. et al. Harnessing human ADAR2 for RNA repair – recoding a PINK1 mutation rescues mitophagy. *Nucleic Acids Res.* **45**, 2797–2808 (2017).
16. Bennett, C. F. et al. Pharmacology of antisense drugs. *Annu. Rev. Pharmacol. Toxicol.* **57**, 81–105 (2017).
17. Heep, M. et al. Applying human ADAR1p110 and ADAR1p150 for site-directed RNA editing-G/C substitution stabilizes guideRNAs against editing. *Genes (Basel)* **8**, 34 (2017).
18. Tan, M. H. et al. Dynamic landscape and regulation of RNA editing in mammals. *Nature* **550**, 249–254 (2017).
19. Lundstrom, K. & Boulikas, T. Viral vectors in gene therapy: technology development and clinical trials. *Technol. Cancer Res. Treat.* **2**, 471–486 (2003).
20. Conrad, N. K. The emerging role of triple helices in RNA biology. *Wiley Interdiscip. Rev. RNA* **5**, 15–29 (2014).
21. Tomatsu, S. et al. Mutations and polymorphisms in GUSB gene in mucopolysaccharidosis VII (Sly Syndrome). *Hum. Mutat.* **30**, 511–519 (2009).
22. BasuRay, S. et al. Rab7 mutants associated with Charcot–Marie–Tooth disease cause delayed growth factor receptor transport and altered endosomal and nuclear signaling. *J. Biol. Chem.* **288**, 1135–1149 (2013).
23. Vogel, P., Hanswillemenke, A. & Stafforst, T. Switching protein localization by site-directed RNA editing under control of light. *ACS Synth. Biol.* **6**, 1642–1649 (2017).
24. McCaffrey, A. P. et al. RNA interference in adult mice. *Nature* **418**, 38–39 (2002).
25. Vabret, N., Bhardwaj, N. & Greenbaum, B. D. Sequence-specific sensing of nucleic acids. *Trends Immunol.* **38**, 53–65 (2017).
26. Wang, D., Tai, P. W. L. & Gao, G. Adeno-associated virus vector as a platform for gene therapy delivery. *Nat. Rev. Drug Discov.* **18**, 358–378 (2019).

Methods

The recruitment cluster finder python tool. The recruitment cluster finder (RCF) tool was programmed in python (v.3.7.2). Data processing of our RCF tool is explained conceptually in Supplementary Fig. 1. The tool allows computation of CLUSTER gRNAs that meet certain input criteria. The input values/strings include the binding site of the specificity domain, the ADAR recruitment motif, the number of RS, their range of length variation and the range of distance variation in RS binding sites on the mRNA. To minimize bystander editing at the RS binding sites, the RCF scans the input DNA, corresponding to the target pre-mRNA/mRNA, for patches of uninterrupted sequences containing only guanosine, thymidine or cytosine, but not adenosine, except for adenosine with 5′ neighboring guanosine. Depending on the input values for the permitted ranges of RS length and distance of RS binding sites on the target mRNA, the RCF selects RS within the determined adenosine-depleted sequence patches. This is followed by brute force assembly and folding (Vienna RNA package[27]) of RS, the specificity domain and the ADAR recruiting domain. The resulting CLUSTER gRNAs are then scored for minimal secondary structure within the antisense part (all RS + linker + specificity domain) by their dot/bracket ratio (based on the dot/bracket notation output generated by Vienna RNA). CLUSTER gRNAs with a high score are less likely to engage in unproductive folding of the antisense part or misfolding of the ADAR recruiting domain. The RS of each gRNA used in the study were analyzed in silico for potential transcriptome-wide off-targets using NCBI blast, which could be automated in the future.

Vector design (cDNA and gRNA). Guide RNA inserts, including the necessary overhangs, were created by hybridization and phosphorylation of oligonucleotides and then cloned into a Hind-III and Bbs-I (CLUSTER gRNA)- or Hind-III and BamHI (LEAPER gRNA)-digested pSilencer 2.1 U6 hygro backbone. Dual-luciferase reporters were created via Gibson cloning of PCR fragments into both a pShuttle (cytomegalovirus (CMV) promoter, SV40 polyA signal, see pTS554 and pTS555) and a pcDNA3.1 (CMV promoter, BGH polyA signal, see pTS656 and pTS657) backbone. Disease-relevant cDNA constructs were created by gene synthesis (ThermoFisher, GeneArt gene synthesis) and cloned into either a pcDNA3.1 backbone (AHI W725X, BMPR2 W298X, COL3A1 W1278X, FANCC W506X, MYBPC3 W1098X, IL2RG W237X) or a pcDNA6 v5-His (PINK1 W437X) via Nhe-I and Xho-I, in both cases under the control of a CMV promoter and terminated by a BGH polyA signal. The sequences of all cloned products were verified by Sanger sequencing. Further details can be found in Supplementary information.

Analysis of RNA editing. Adenosine-to-inosine editing yields were quantified from Sanger sequence traces, based on the relative height of the signal of guanosine compared to the sum of guanosine plus adenosine, as described previously[15]. If a reverse primer was used for sequencing, cytidine and thymidine peaks were treated accordingly. For better comparability, on-target editing yields of the same target with different gRNAs were quantified using the same sequencing primer. Off-target editing had to be evaluated with different sequencing primers in most cases, due to large distances between gRNA binding sites. Only the cleanest reads were used for off-target evaluations, whereas guanosine peaks below background were counted as 0% off-target. Editing events with yields <10% were background corrected with the negative control.

Reverse transcription quantitative PCR experiments. RNA isolation was performed using either the High Pure miRNA Isolation Kit from Roche (using a binding enhancer) or the PARIS Kit from Invitrogen (300 µl of fractionation and disruption buffer, 2× lysis/binding solution at room temperature, an optional washing step with 300 µl of fractionation buffer, centrifugation of fractionated samples for 5 min at 4 °C and 500*g*, elution from filter cartridges first with 40 µl, followed by 10 µl of elution-buffer heated to 95 °C), both according to the manufacturer's protocol. The High Pure miRNA Isolation Kit was used for quantification experiments (Supplementary Fig. 6) while the PARIS Kit was used for localization experiments (Supplementary Fig. 9). DNase digestion was performed according to the manufacturer's protocol (rigorous two-step incubation treatment) using the Turbo DNase Kit (ThermoFisher, no. AM1907). Reverse transcription with 500 ng of RNA per sample was performed using the high-capacity cDNA reverse transcription kit (ThermoFisher, no. 4368814), followed by PCR clean-up using the NucleoSpin Gel and PCR clean-up kit (Macherey Nagel, no. 740609). Quantitative PCR (qPCR) was executed in an Applied Biosystems 7500 qPCR machine (96-well qPCR-plate, 20 ng of cDNA (10 ng µl^{-1}) per well). The Fast Sybr-Green mastermix (Applied Biosystems, no. 4385612) was used according to the manufacturer's protocol (10 µl of Sybr-Green-Mix, 7.2 µl of nuclease-free water and 0.4 µl of each primer). Samples were measured in two to three technical replicates, while TE-buffer-negative controls were measured in duplicate. The run method is shown in Supplementary Fig. 13. A baseline correction was performed for each dataset using the 7500 data analysis software before C(t) values were acquired. ΔΔC(t)-calculations were performed as further described in Supplementary Fig. 14 (refs. [28,29]). Amplification efficiency and melting curves were analyzed for each primer pair using a cDNA dilution series (Supplementary Figs. 15 and 16). To normalize gRNA data in quantification experiments, U6 small nuclear RNA was used as the reference gene. For quantification and localization experiments, U6 snRNA and Malat1 were used as nuclear reference genes with GAPDH and HPRT1 as cytoplasmic genes.

Immunoblotting. For immunoblotting, cells were harvested and lysed in urea lysis buffer (8 M urea, 100 mM NaH_2PO_4, 10 mM Tris, pH 8.0). Shear force was applied using a 23-gauge syringe, and cell debris was removed by centrifugation at 30.000*g* for 15 min at 4 °C. A Bradford assay was then used to normalize total protein amounts, and appropriate amounts of protein lysate in 1× Laemmli buffer were loaded for SDS–polyacrylamide gel electrophoresis (Novex 8–16% Tris-Glycine Mini Gel; ThermoFisher, no. XP08165BOX). Proteins were transferred onto a polyvinylidene difluoride membrane using a tank-blotting system at 35 V overnight. The membrane was blocked in 5% nonfat dry milk Tris buffered saline with Tween (TBS-T) for 1 h at room temperature, and afterwards incubated with the primary antibodies (5% nonfat dry milk TBS-T plus 1:1,000 anti-ADAR1, Santa Cruz, no. sc-73408, plus 1:5,000 anti-β-actin, Sigma-Aldrich, no. A5441) at 4 °C overnight. Secondary antibodies (5% nonfat dry milk TBS-T plus 1:10,000 anti-mouse horseradish peroxidase; Jackson ImmunoResearch Laboratories, no. 115-035-003) were incubated for 1.5 h at room temperature. After each antibody incubation, the membrane was washed three times for 5 min with TBS-T. Detection was performed using 1 ml of Clarity Western ECL Substrate (Bio-Rad) and an Odyssey Fc imaging system (LI-COR). For further details on antibodies, see Supplementary Table 5. Results are reported in Supplementary Figs. 3 and 12.

Cell culture. In general, cell cultures were grown in DMEM (ThermoFisher, no. 41965062) supplemented with 10% fetal bovine serum (FBS; ThermoFisher, no. 10270106) for cell lines and 15% FBS for primary fibroblasts. All cell cultures were kept in an incubator at 37 °C and 5% CO_2.

Editing readout via Sanger sequencing (total RNA from cell lines). Cells were harvested in RLT buffer (Qiagen, no. 79216) followed by RNA isolation using either the Monarch RNA cleanup kit (NEB, no. T2030L) or the RNeasy Mini RNA isolation kit (Qiagen, no. 74104). DNase-I digestion was performed according to the manufacturer's manual using DNase-I (NEB, no. M0303S). One-step PCR with reverse transcription (RT–PCR) was performed using the Biotechrabbit One-Step RT-PCR Kit (Biotechrabbit, no. BR0400102) for regular substrates. For difficult substrates (for example, PDE4D L291L), the OneTaq One-Step RT–PCR Kit (NEB, no. E5315S) was used. Immediately before RT–PCR, samples were mixed with the respective sense oligo (1 µl of 10 µM) corresponding to the gRNAs used (Supplementary Table 4) and heated to 90 °C for 2 min. After TAE-agarose gel electrophoresis and PCR cleanup (NucleoSpin Gel and PCR Clean-up Kit, Macherey Nagel, no. 740609), Sanger sequencing (Microsynth AG) was performed.

Editing readout via Sanger sequencing (total RNA from primary fibroblasts). Cells were harvested in RLT buffer (Qiagen, no. 79216), followed by RNA isolation using the Monarch RNA cleanup kit (NEB, no. T2030L). Turbo DNase digestion was performed using the Turbo DNA-free Kit (ThermoFisher, no. AM1907). Reverse transcription was performed using the SuperScript IV RT (ThermoFisher, no. 18090050) with random primers for 1 h at 52 °C. A regular PCR followed by nested PCR was performed using Taq Polymerase (NEB, no. M0267S), with each reaction containing 10% DMSO. After the second PCR the products were separated by SB-agarose gel electrophoresis. Following PCR cleanup (NucleoSpin Gel and PCR Clean-up Kit, Macherey Nagel, no. 740609), Sanger sequencing (Microsynth AG) was performed.

Editing readout via dual-luciferase activity. Dual-luciferase activity was measured with the Dual-Luciferase Reporter Assay System (Promega) according to the manual, in 96-well scale. Cells were washed with PBS, then lysed in 1× passive lysis buffer (35 µl per well) while shaking for 15 min at 700 r.p.m. at room temperature. Cell lysate (30 µl per well) was transferred to a LumiNunc 96-well plate (VWR, no. 732-2696) and measured in a Spark 10 M plate reader (Tecan) using the dual-luciferase reporter assay reagents (35 µl per well) with an auto-injector. For plasmid transfection (lipofection), Tecan reader standard settings were used (OD-None). To prevent overload of the sensor following viral delivery, signal reduction was applied (OD-1). Each measurement was performed for 10 s, starting 5 s after injection. Per treatment, five biological replicates were analyzed, each measured in one technical replicate. For data processing, measured blank values (background) were subtracted from samples and controls then all firefly values were divided by the corresponding renilla values. The resulting normalized firefly activity of all samples was then set in ratio to either the positive control, to obtain the restored normalized firefly activity as a percentage, or to a specific sample, to report firefly activity as fold change relative to that sample.

Editing readout via α-L-iduronidase enzyme activity assay. A standard dilution series of 4-methylumbelliferone (Sigma-Aldrich, no. M1381) was prepared in 1× PBS. For each concentration, 25 µl of the standard solution was added to 25 µl of 0.4 M sodium formate buffer (pH 3.5) and applied to a 96-well LumiNunc plate (VWR, no. 732-2696) in triplicate. The substrate (4-methylumbelliferyl α-L-iduronide, Glycosynth, no. 44076) was dissolved in 0.4 M sodium formate

buffer to a final concentration of 180 µM. For the murine IDUA assay using HeLa cells, 25 µl of 1:3 diluted cell lysate (0.5% Tween-20/PBS) was added to 25 µl of substrate in the plate and incubated for 45 min at 37 °C in the dark. For the human IDUA assay using primary fibroblasts, 25 µl of undiluted cell lysate (0.5% Triton X-100/PBS) was added to 25 µl of substrate in the plate and incubated for 90 min at 37 °C in the dark. The reaction was quenched in both cases by the addition of 200 µl of glycine carbonate buffer (0.17 M glycine/NaOH, pH 10.4). The fluorescence of 4-methylumbelliferone was measured at an excitation wavelength of 355 nm and an emission wavelength of 460 nm with a Tecan Spark 10 M plate reader. Calculated enzyme activities were referenced to the protein amount as determined by bicinchoninic acid (BCA) assay (Pierce BCA Protein Assay Kit, ThermoFisher, no. 23227). For the murine IDUA assay in HeLa cells, enzyme activity was standardized to HeLa cells transfected with wild-type murine IDUA plasmid. For the human IDUA assay in primary fibroblasts, enzyme activity was standardized to Scheie fibroblast lysate. The results are reported in Fig. 3b,h.

HEK293FT cells (culture settings for Sanger sequencing). HEK293FT cells (6×10^4) were seeded in 24-well scale in 450 µl of DMEM and 10% FBS. After 24 h, cells were transfected with 1,200 ng of gRNA plasmid (NucleoSpin Plasmid Transfection-grade, Macherey Nagel, no. 740490) using a 1:3 ratio of FuGene6 (Promega, no. E2691). Forty-eight hours after transfection, cells were harvested. As the readout method, Sanger sequencing was used. Results are reported in Figs. 3c–e and 4.

ADAR-Flp-In T-REx cells (culture settings for Sanger sequencing). Either 2.5×10^5 ADAR1 p110, p150 Flp-In T-REx cells or 3×10^5 ADAR2 Flp-In T-REx cells were seeded on poly-D-lysine (PDL)-coated 24-well plates in 500 µl of DMEM, 10% FBS and 10 ng ml^{-1} doxycycline. After 24 h, cells were transfected with 300 ng of dual-luciferase reporter and 1,300 ng of gRNA plasmid (NucleoSpin Plasmid Transfection-grade, Macherey Nagel, no. 740490) using a 1:3 ratio of Lipofectamine-2000 (ThermoFisher, no. 11668019). Cells were harvested 72 h after transfection. As the readout method, Sanger sequencing was used. Results are reported in Fig. 1c,d.

ADAR-Flp-In T-REx cells (culture settings for dual-luciferase assay). Either 4×10^4 ADAR1 p110 or p150 Flp-In T-REx cells were seeded on PDL-coated 96-well plates in 100 µl of DMEM, 10% FBS and 10 ng ml^{-1} doxycycline. After 24 h, cells were transfected with 60 ng of dual-luciferase reporter and 260 ng of gRNA plasmid (NucleoSpin Plasmid Transfection-grade, Macherey Nagel, no. 740490) using a 1:0.8 ratio of Lipofectamine-2000 (ThermoFisher, no. 11668019). The luciferase assay was performed 72 h post transfection. Results are reported in Fig. 1b.

HeLa cells (culture settings for Sanger sequencing). HeLa cells (1.2×10^5) were seeded in 24-well scale in 500 µl of DMEM and 10% FBS. Cells were transfected 24 h after seeding with 800 ng gRNA plasmid and 200 ng of target-encoding plasmid (NucleoSpin Plasmid Transfection-grade, Macherey Nagel, no. 740490) per well using a plasmid:Lipofectamine-3000 ratio of 1:1.5. Cells were harvested 72 h after transfection. As the readout, Sanger sequencing was used. Results are reported in Figs. 1e, 3a,b and 5b.

HeLa cells (culture settings for dual-luciferase assay). HeLa cells (2.4×10^4) were seeded in 96-well scale. Cells were transfected 24 h after seeding with 160 ng of gRNA plasmid and 40 ng of dual-luciferase reporter per well using a plasmid:Lipofectamine-3000 ratio of 1:1.5 and a plasmid:P3000 reagent ratio of 1:2. The luciferase assay was performed 48 h post transfection. Results are reported in Figs. 2a–f and 5b.

HeLa cells (culture settings for murine α-L-iduronidase enzyme activity assay). HeLa cells (1.0×10^5) were seeded in 24-well scale in 500 µl of DMEM and 10% FBS. Cells were transfected 24 h after seeding with 800 ng of gRNA plasmid and 200 ng target-encoding plasmid (NucleoSpin Plasmid Transfection-grade, Macherey Nagel, no. 740490) per well using a plasmid:Lipofectamine-3000 ratio of 1:1.5. Cells were harvested 72 h after transfection in 160 µl 0.5% Tween-20/PBS and incubated on ice for 30 min. After 1:3 dilution in 0.5% Tween-20/PBS, the BCA and murine α-L-iduronidase enzyme activity assays were performed as described.

HeLa cells (culture settings for siRNA knockdown of ADAR isoforms and editing). HeLa cells were reverse transfected in 12-well format with 2.5 pmol of siRNA against ADAR1 (both isoforms, Dharmacon, SMARTpool: ON-TARGETplus ADAR (103) siRNA, L-008630-00-0005), ADAR1 p150 (Ambion (Life Technologies), sense strand: 5′-GCCUCGCGGGCGCAAUGAAtt; antisense strand: 5′-UUCAUUGCGCCCGCGAGGCat) or mock (Dharmacon, siGENOME Non-Targeting siRNA Pool no. 2, D-001206-14-05). For this reaction, 200 µl of transfection mix, containing 2.5 µl of the respective siRNA (10 nM), 3 µl of HiPerFect (Qiagen) and Opti-MEM, was distributed evenly in each well before the addition of 800 µl of DMEM and 10% FBS containing 1.2×10^5 HeLa cells. Medium was changed every 24 h. For RNA-editing experiments, cells were detached 24 h after siRNA transfection and reseeded in 96-well scale at a density of 5×10^4 cells per well; 24 h later the cells were transfected using 160 ng of gRNA plasmid and 40 ng of dual-luciferase reporter per well, at a plasmid:Lipofectamine-3000 ratio of 1:1.5 and a plasmid:P3000 reagent ratio of 1:2. The luciferase assay was performed 48 h post transfection. Results are reported in Fig. 1f.

Human primary fibroblasts (culture settings for human α-L-iduronidase enzyme activity assay and Sanger sequencing). Fibroblasts from patients with Scheie syndrome (GM01323) and Hurler syndrome (GM06214) were purchased from the Coriell Institute for Medical Research. Cells (2.5×10^5 per well in 2.5 ml of DMEM and 15% FBS) were seeded in six-well plates. For each tested condition, two six-well sections were used for the IDUA assay and one was used to determine RNA-editing yields by Sanger sequencing. The CLUSTER ASO was a polyacrylamide gel electrophoresis-purified, endblocked (2′-OMe, PS) RNA oligonucleotide with a 3×RS (20-20p8-25-20) CLUSTER design, which was ligated (T4 RNA ligase) in-house from two commercially purchased (Biospring) and high-performance liquid chromatography-purified oligonucleotides of length 69 nt (5′ part) and 80 nt (3′ part), according to our recently published protocol[30]. The full sequence and modification pattern is given in Supplementary Table 1). Transfection was performed 24 h after seeding with 125 pmol of ASO and 7.5 µl of RNAiMAX, each diluted in 250 µl of Opti-MEM. Both solutions were combined after 5 min of incubation and incubated for an additional 20 min before the transfection mix was distributed evenly into one well. The medium was changed 24 h after transfection; 48 h after transfection, fibroblasts were detached and washed once with PBS, 40 µl of 0.5% Triton X-100 in PBS was added to the cell pellet and incubated on ice for 30 min and α-L-iduronidase enzyme assay was then performed. Results are reported in Fig. 3g,h.

Production of adenoviral vectors. Dual-luciferase constructs were cloned into the pShuttle-CMV backbone (Addgene, no. 16403), while gRNA constructs were cloned into the pShuttle backbone (Addgene, no. 16402). Pme-I (NEB, no. R0560S) linearized shuttle vectors containing the gene of interest were then delivered into BJ5183-AD-1 *Escherichia coli* (Agilent, no. 200157) via electroporation with a Bio-Rad Genpulser at 1.6 kV, 200 Ohm and 25 µF. Plasmids were isolated from bacterial cultures using a Gravity-Flow Plasmid Mini-Kit (Qiagen, no. 12123). Recombinant adenoviral plasmids were verified as containing the gene of interest by Pac-I (NEB, no. R0547S) control digestion. For larger-scale production of recombined plasmids, these were retransformed and isolated from bacterial cultures using a Gravity-Flow Plasmid Midi-Kit (Qiagen, no. 12143). For virus production, 30–100 µg of adenoviral production plasmid was digested with Pac-I and purified by ethanol precipitation; 10 µg of the Pac-I-digested adenoviral production plasmid was then transfected into 40–80% confluent Ad293 cells (15-cm plate) using a plasmid:Lipofectamine-2000 ratio of 1:3. Within 7–10 days the emerging widespread cytopathic effect indicated successful adenovirus production. Cells were then harvested, pelletized and treated with three freeze–thaw cycles at −80 °C in 1 ml of PBS to release adenovirus. Lysates were cleared from cell debris by centrifugation (10 min, 3,000 r.p.m.) and stored for further use at −80 °C after the addition of 10% glycerol. After two further rounds of amplification, the virus was purified by CsCl density-gradient ultracentrifugation for 2 h at 32,000 r.p.m. (layer composition: 3 ml of 1.41 g ml^{-1} CsCl in PBS at the bottom, 4.5 ml of 1.27 g ml^{-1} CsCl in PBS in the middle and ~3 ml of crude adenovirus in PBS/glycerol on top). Adenovirus particles were extracted from the space between the 1.41 and the 1.27 g ml^{-1} CsCl layers. A further ultracentrifugation step, using a bottom layer with 10 ml of 1.34 g ml^{-1} CsCl in PBS and a top layer of the retrieved adenovirus, was followed immediately by ultracentrifugation at 32,000 r.p.m. for 18–24 h. The adenovirus layer was extracted again and inserted into a dialysis cassette (10 K Slide-A-Lyzer Dialyis Cassette G2, ThermoFisher, no. 87730). Dialysis was performed in 1 l of dialysis buffer (10 mM Tris-base, 0.5% glycerol, pH 8.0) at 4 °C overnight under constant stirring (magnetic stir bar) and buffer exchange at 1, 2, 3, 5 and 7 h. After retrieval, 10% glycerol was added. The adenovirus was then aliquoted and frozen at −80 °C. Titer was determined using the Adeno-X Rapid Titer Kit (CloneTech/Takara, no. 632250) performed according to the manufacturer's protocol. Purified viruses were tested for E1A negativity, and thus replication deficiency, by immunoblot (Mouse Anti-Adenovirus Type 5 E1A, BD Pharmingen, no. 554155).

Adenovirus transduction of a panel of cell lines (culture settings for dual-luciferase assay and Sanger sequencing). A total of 2.5×10^4 cells per well were reverse infected with gRNA and dual-luciferase reporter adenoviruses (96-well scale, in DMEM and 10% FBS). The cell lines used were A549 (European Collection of Authenticated Cell Cultures, no. ECACC 86012804), HeLa cells (catalog no. ATCC CCL-2), HepG2 (DSMZ catalog no. ACC180), Huh7 (CLS, catalog no. 300156), SH-SY5Y (catalog no. ATCC CRL-2266), SK-N-BE(2) (catalog no. ATCC CRL-2271) and U87MG (catalog no. ATCC HTB-14). Cell lines deriving from cervix (HeLa), liver (Huh7, HepG2) and lung (A549) were infected with 75, 100 or 125 MOI of gRNA adenovirus and with 50 MOI of dual-luciferase WT/amb reporter adenovirus (wild-type renilla luciferase, W417X amber firefly luciferase). Neuroblastoma cell lines (SK-N-BE(2), SH-SY5Y and U87 MG) were reverse infected with 100, 250 or 500 MOI of gRNA adenovirus and 50 MOI of dual-luciferase WT/amb reporter. Negative control cells were infected with 50 MOI of dual-luciferase WT/amb adenovirus. Positive control cells were infected

with 5 MOI of dual-luciferase WT/WT adenovirus (wild-type renilla luciferase, wild-type firefly luciferase). Medium was changed daily; 96 h post transfection, cells were harvested in either 1× passive lysis buffer for dual-luciferase assay readout or RLT buffer for Sanger sequencing readout. The results are reported in Fig. 1g, h.

Next-generation RNA-seq experiment. The RNA-editing experiment was done by transfection of 1,200 ng of gRNA plasmid (NucleoSpin Plasmid Transfection-grade, Macherey Nagel, no. 740490) into 6×10^4 HEK293FT cells 24 h post seeding using a 1:3 ratio of FuGene6 (Promega, no. E2691) in 24-cell format; 48 h after transfection, cells were harvested. Overall, three settings were carried out, each with an independent duplicate: (1) nontargeting gRNA (NT-RNA), (2) RAB7A 3′-UTR 19-11-13-20p8 CLUSTER gRNA and (3) RAB7A 3′-UTR 111p56 LEAPER gRNA. RNA was isolated with the RNeasy MinElute Kit (Qiagen, no. 74204), treated with DNase-I (NEB, no. M0303S), incubated with an RNA strand reverse complementary to the antisense part of the respective gRNA, heated to 95 °C for 3 min and purified again with the RNeasy MinElute Kit. Purified RNA was delivered to CeGaT for poly(A) + mRNA-seq. The library was prepared from 200 ng of RNA with the TruSeq Stranded mRNA Library Prep Kit (Illumina) and sequenced with a NovaSeq 6000 (50 M reads, 2 × 100-bp paired-end; Illumina). Results are reported in Fig. 4c–f and Extended Data Figs 6, 7, 8 and 9.

Mapping of RNA-seq and reads. We adopted a previously published pipeline to accurately align RNA-seq reads onto the genome[31,32]. We used STAR (v.2.5.3a)[33] to align reads to the hg19 reference genome and ran Picard tools (v.1.129) to remove clonal reads (PCR duplicates) mapped to the same location. Of these identical reads, only the read with the highest mapping quality was kept for downstream analysis. Unique and nonduplicate reads were subjected to local realignment and base score recalibration using the IndelRealigner and TableRecalibration from the Genome Analysis Toolkit (GATK, v.3.6)[34]. The above steps were applied separately to each RNA-seq sample.

Identification of editing sites from RNA-seq data. To remove LEAPER gRNA sequences falsely aligned to the targeting region, we used the rmdup command in samtools[34] to remove PCR duplicates in the RAB7A 3′-UTR region (Extended Data Fig. 10). Additionally we removed all reads containing the sequence 'AAGGGTG' (3′ end of LEAPER gRNA) and those that ended with 'TCAAAGAC' (5′ end of LEAPER gRNA). As a final step we removed all reads that originated from the antisense sequence of the RAB7A gene. This procedure was applied to all samples (CLUSTER, LEAPER and gNT-RNA). To call variants from the mapped RNA-seq reads we used the UnifiedGenotyper from GATK[35]. In contrast to the usual practice of variant calling, we identified those variants with relatively loose criteria using the UnifiedGenotyper tool with options stand_call_conf 0, stand_emit_conf 0 and output mode EMIT_VARIANTS_ONLY. Variants from nonrepetitive and repetitive non-Alu regions were required to be supported by at least three reads containing mismatches between the reference genome sequences and RNA-seq. Supporting of one mismatch read was required for variants in Alu regions. This set of variant candidates was subjected to several filtering steps to increase the accuracy of editing site calling. We first removed all known human single-nucleotide polymorphism (SNPs) present in dbSNP build 137 (except SNPs of molecular type 'cDNA'; database v.135; http://www.ncbi.nlm.nih.gov/SNP/), the 1000 Genomes Project and the University of Washington Exome Sequencing Project (http://evs.gs.washington.edu/EVS/). To remove false-positive RNA-seq variant calls due to technical artifacts, further filters were applied as previously described[30,31]. In brief, we required a variant call quality $Q>20$ (refs. [30,31]), discarded variants if they occurred in the first six bases of a read, removed variants in simple repeats[34], removed intronic variants that were within 4 bp of splice junctions and discarded variants in homopolymers. Moreover, we removed reads mapped to highly similar regions of the transcriptome by BLAT[36]. Finally, variants were annotated using ANNOVAR[37] based on gene models from Gencode, RefSeq, Ensembl and UCSC. All sites identified from RNA-seq data were compared with all sites available in the RADAR database[38], referred as 'known' sites if found in RADAR or 'novel' sites if not.

Identification of significantly differently edited sites. We merged all sites found in the RNA-seq samples with the sites of the RADAR database, quantified editing levels of edited sites with ≥50 reads coverage (combined coverage of both replicates) and performed Fisher's exact tests followed by Benjamini–Hochberg multiple test correction (adjusted $P<0.01$), to identify significantly differently edited sites across samples (absolute editing difference >10%).

Measurement of RAB7A editing precision. To compare the specificity of CLUSTER and LEAPER gRNA, we selected all mapped reads containing the edited target sequence GCTGGCGG. Edited reads and their partner reads were compared to the RAB7A sequence covering the editing region to identify A-to-G mismatches. As a control, we used the nontargeting sample to quantify A-to-G mismatches in reads covering the unedited target sequence GCTAGCGG.

Gene expression analysis. We ran RSEM (v.1.2.21)[39] on STAR alignments to obtain read counts and transcripts per million (TPM) values to characterize gene expression. We used read counts to perform expression analysis with DESeq[40] for all genes with TPM ≥2 (for both replicates) and identified significantly expressed genes using a threshold of $P_{adj}<0.01$ and $|\log_2$ fold change$|>2$.

Animal experiments. *Laboratory animals.* Mice, strain C57BL/6, age 5–6 weeks, were split into groups. The negative control, positive control and the editing group 2 (20-15-15-20p8 gRNA treatment) comprised equal numbers of male and female animals; editing group 1 (15-13-11-20p8 gRNA treatment) consisted of females only.

Handling. All C57BL/6 mice were kept under specific-pathogen-free conditions and controlled humidity (target value 55%, minimum 45%, maximum 65%), temperature (target value 22 °C, minimum 20 °C, maximum 24 °C) and lighting (12/12-h light/dark period) and free access to food (regular mouse chow) and water. Mouse experiments complied with German animal experimentation regulations and were approved by the Ethics Review Committee of the regional council in Tübingen (reference no. 35/9185.81-2 / M 2/18).

Hydrodynamic tail vein injection. Mice were treated with endotoxin-free plasmids (NucleoBond Xtra Midi EF, Macherey Nagel) diluted in saline solution in a total volume equal to 10% of their body weight. Hydrodynamic tail vein injection of the total volume was performed within 5–10 s using a 25½-gauge needle. Negative control group mice were treated with 10 µg of dual-luciferase WT/amb reporter plasmid, positive control group mice with 10 µg of dual-luciferase WT/WT reporter plasmid and editing group mice with 5 µg of dual-luciferase WT/amb reporter plasmid and 25 µg of gRNA plasmid. The mice were sacrificed 72 h after injection. At this time all animals appeared to be completely healthy and did not show any stress symptoms. The liver was removed, the lobes separated and each cut in three. One piece was pooled with pieces of other lobes from the same animal and then used for the dual-luciferase assay. The two remaining pieces of each liver lobe were placed in separate Eppendorf tubes and immediately frozen in liquid nitrogen, one of which was then used for RNA isolation.

Dual-luciferase assay. Dual-luciferase activity was measured with the Dual-Luciferase Reporter Assay System (Promega) according to the manual. For this, pooled liver lobe pieces were homogenized in passive lysis buffer (500 µl) using a micropestle. After spinning debris down, 50 µl of sample per well was transferred onto a white, 96-well LumiNunc plate (VWR, no. 732-2696). Each sample was measured in triplicate with a Tecan Spark 10 M plate reader equipped with an auto-injector, using 35 µl of each assay substrate per well. For data processing, measured blank values (background) were subtracted from samples and controls and then all firefly values were divided by the corresponding renilla values. The resulting normalized firefly activity of all samples was then set in ratio to the positive control, to obtain the restored normalized firefly activity as a percentage. Under each condition, $n=3$–5 mice were used. From each mouse, luciferase activity was determined in each of the five liver lobes individually in technical triplicate. In Fig. 5a, each data point represents one mouse reporting the mean of fifteen measurements.

RNA isolation and Sanger sequencing. RNA isolation from the liver was done with TRIzol reagent (ThermoFisher, no. 15596026) according to the manual. Pieces of individual liver lobe were homogenized separately for each of the five lobes per animal in TRIzol reagent (1.0 ml) with a micropestle. Chloroform (200 µl) was added and vortexed. After incubation at room temperature (5–10 min), samples were centrifuged at 4 °C (12,000*g*, 20 min). Isopropanol (700 µl) was added to the aqueous phase for precipitation overnight at −20 °C. Precipitated RNA was centrifuged (14,000*g*, 60 min), washed twice with 75% EtOH (500 µl), dried (50 °C, 3 min) and dissolved in nuclease-free water (87.5 µl). Then, RNA was DNase-I digested (30 min, 37 °C; DNase-I buffer, 2.5 µl of DNase-I, NEB, no. M0303S). RNA was cleaned up using the RNeasy Mini RNA Isolation Kit (Qiagen, no. 74104). To remove residual plasmid DNA, a second DNase digestion was required. For this, Turbo DNase (ThermoFisher, no. AM1907) was used according to the established protocol (rigorous two-step incubation treatment). Reverse transcription was performed using ProtoScript II reverse transcriptase (NEB, no. M0368S) and random primer mix (High-Capacity cDNA Reverse Transcription Kits, Applied Biosystems, no. 4368814) with 1 µg of total RNA. After PCR cleanup (NucleoSpin Gel and PCR Clean-up Kit, Macherey Nagel, no. 740609), 2.5 µl of cDNA was used for the Q5-polymerase (NEB, no. M0491S) PCR using primer pair 2898 and 2899, followed by nested PCR using Taq-polymerase (NEB) and primer pair 3850 and 3851. Sanger sequencing (Microsynth AG) was performed using primer 3850. Under each condition, $n=3$–5 mice were used. From each mouse, RNA-editing yield was determined in each of the five liver lobes individually (Supplementary Fig. 10). In Fig. 5a, each data point represents one mouse reporting the mean of five measurements.

Data analysis. Non-NGS data were analyzed using Excel 2016 and GraphPad Prism 8. Figures were created with CorelDraw 2017. The manuscript was written with Word 2016 and the custom 'recruitment cluster finder' tool was written in python 3.7.2 using the Vienna RNA package 2.0. Guide RNA folds were

created using the nucleic acid package (Caltech, Nupack.org). qPCR analysis was performed using 7500 data analysis software v.2.3.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Transcriptome-wide RNA-seq data are accessible via the NCBI GEO database with accession code GSE184244. The RADAR database (http://RNAedit.com) was used as reference to assess the editing homeostasis. The dbSNP database version 135 (http://www.ncbi.nlm.nih.gov/SNP/) was used to discriminate human SNPs from RNA editing events. NGS artefacts caused by the LEAPER gRNA treatment were confirmed within transcriptome-wide RNA-seq data from Qu et al.[13] (NCBI Sequence Read Archive database, accession code PRJNA544353).

Code availability

A conceptual description of the RCF tool used to generate CLUSTER gRNAs is provided in Supplementary Fig. 1. The full code is available upon request. The compiled tool is available online under: https://github.com/recruitment-cluster-finder/rcf/releases.

References

27. Lorenz, R. et al. ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **6**, 26 (2011).
28. Wang, Y., Zhu, W. & Levy, D. E. Nuclear and cytoplasmic mRNA quantification by SYBR green based real-time RT-PCR. *Methods* **39**, 356–362 (2006).
29. Schmittgen, T. D. & Livak, K. J. Analyzing real-time PCR data by the comparative C(T) method. *Nat. Protoc.* **3**, 1101–1108 (2008).
30. Merkle, T. & Stafforst, T. New frontiers for site-directed RNA editing: harnessing endogenous ADARs. *Methods Mol. Biol.* **2181**, 331–349 (2021).
31. Ramaswami, G. et al. Accurate identification of human Alu and non-Alu RNA editing sites. *Nat. Methods* **9**, 579–581 (2012).
32. Ramaswami, G. et al. Identifying RNA editing sites using RNA sequencing data alone. *Nat. Methods* **10**, 128–132 (2013).
33. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
34. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
35. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
36. Kent, W. J. BLAT–the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
37. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
38. Ramaswami, G. & Li, J. B. RADAR: a rigorously annotated database of A-to-I RNA editing. *Nucleic Acids Res.* **42**, D109–D113 (2014).
39. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
40. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

Acknowledgements

We thank Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) (nos. STA 1053/3-2, STA 1053/7-1, STA 1053/10-1 and STA 1053/11-1 to T.S. and nos. FOR2314 – 267467939, SFB-TR209 – 314905040 and SFB-TR240 to L.Z.); the Gottfried Wilhelm Leibniz Program (to L.Z.); DFG, under Germany's excellence strategy (no. EXC 2180 – 390900677, Image Guided and Functionally Instructed Tumor Therapies, to L.Z. and T.-W.K.); and the Federal Ministry of Education and Research and Baden-Württemberg Ministry of Science as part of the Excellence Strategy of the German Federal and State Governments, with an intramural innovation grant (nos. IG-Reauts-2019-05 (to P.R.) and IG-RNA-Edit (to J.W.)). We thank the International Rett Syndrome Foundation (grant no. 3806 to T.S.). Further funding was provided by the European Research Council ('CholangioConcept' to L.Z. and T.-W.K.) and the German Center for Translational Cancer Research (to L.Z. and T.-W.K.). We thank the German National Academy of Sciences Leopoldina (grant no. LPDS 2019-06 to P.V.).

Author contributions

P.R. and T.S. conceived and analyzed the overall study. P.R. and N.W. conceived and realized software implementation. P.R., N.L., T.-W.K., L.Z. and T.S. conceived, performed and analyzed animal experiments. P.R., J.W., U.N. and T.S. conceived, performed and analyzed adenovirus experiments. P.R., J.W., A.S., C.Z., N.L. and L.S.P. performed all other wet laboratory experiments. P.V. and J.B.L. analyzed, and all authors interpreted, NGS data. P.R. and T.S. wrote the manuscript. All authors proofread the manuscript.

Competing interests

T.S., J.W., P.R., L.S.P., N.L., P.V. and J.B.L. hold patents on site-directed RNA editing. T.S., J.W., P.R. and N.W. are inventors of a filed patent based on the work published here. P.V., J.B.L. and T.S. are cofounders and shareholders of AIRNA Bio.

Additional information

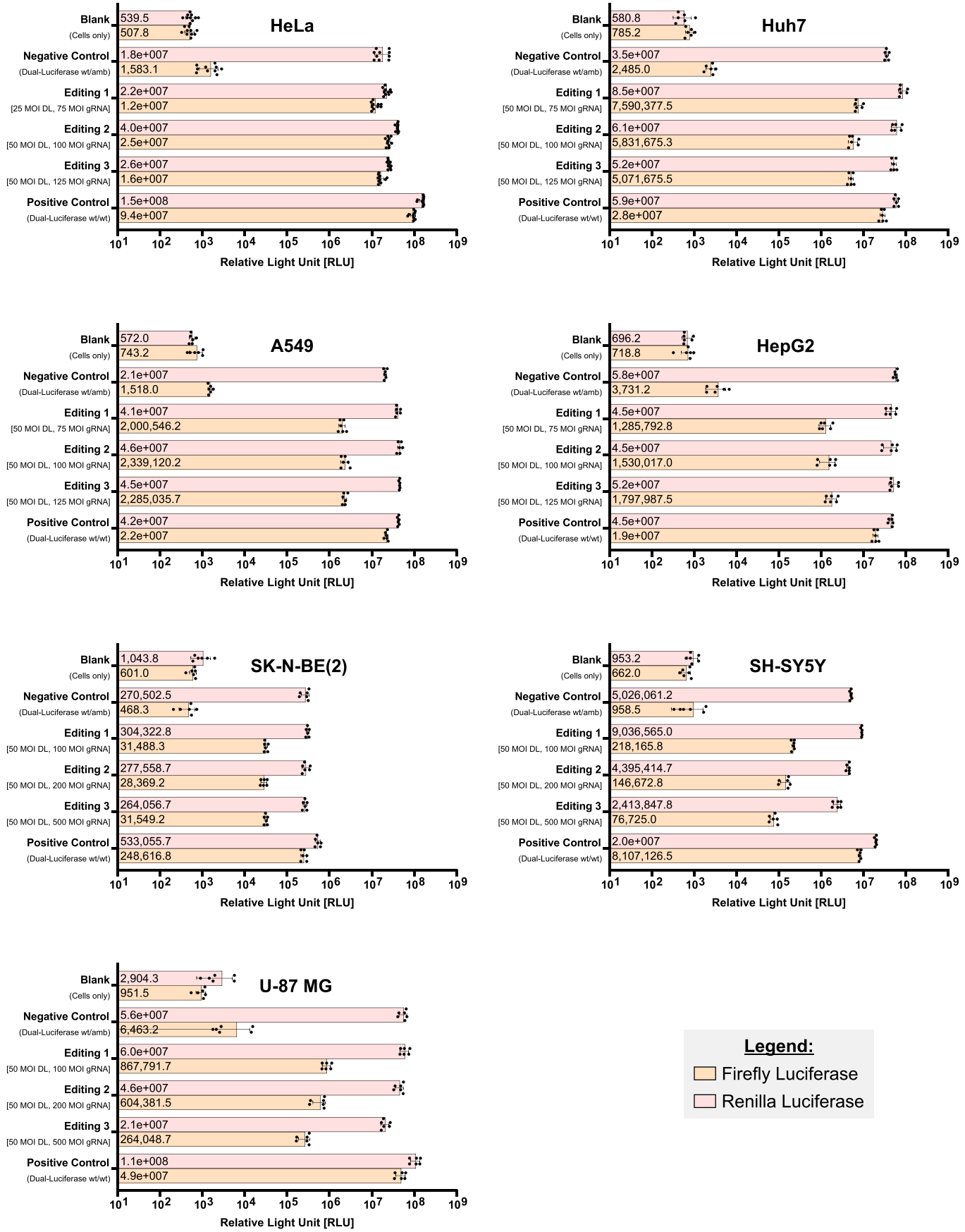
Extended data is available for this paper at https://doi.org/10.1038/s41587-021-01105-0.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41587-021-01105-0.

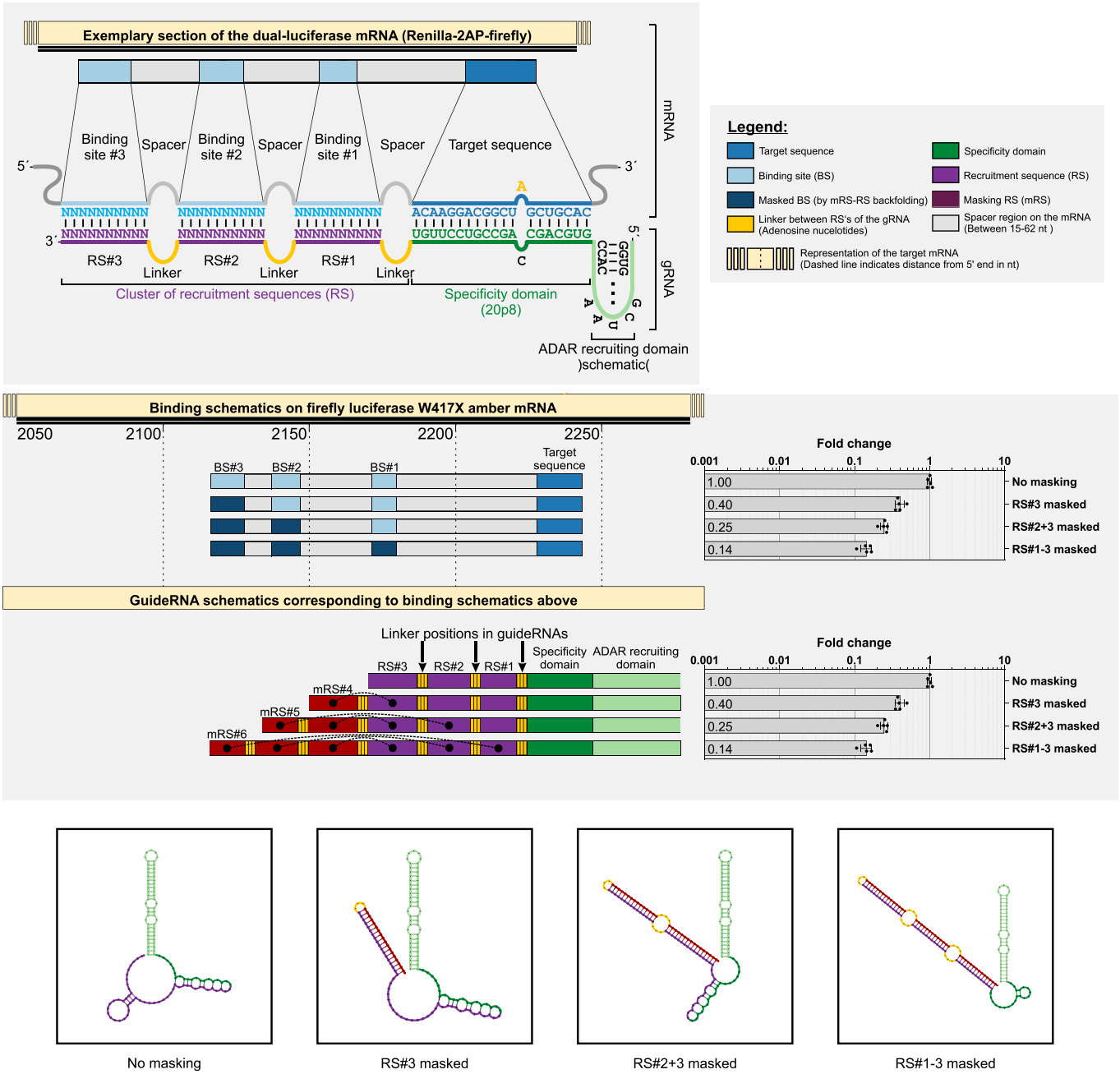
Correspondence and requests for materials should be addressed to Thorsten Stafforst.

Peer review information *Nature Biotechnology* thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

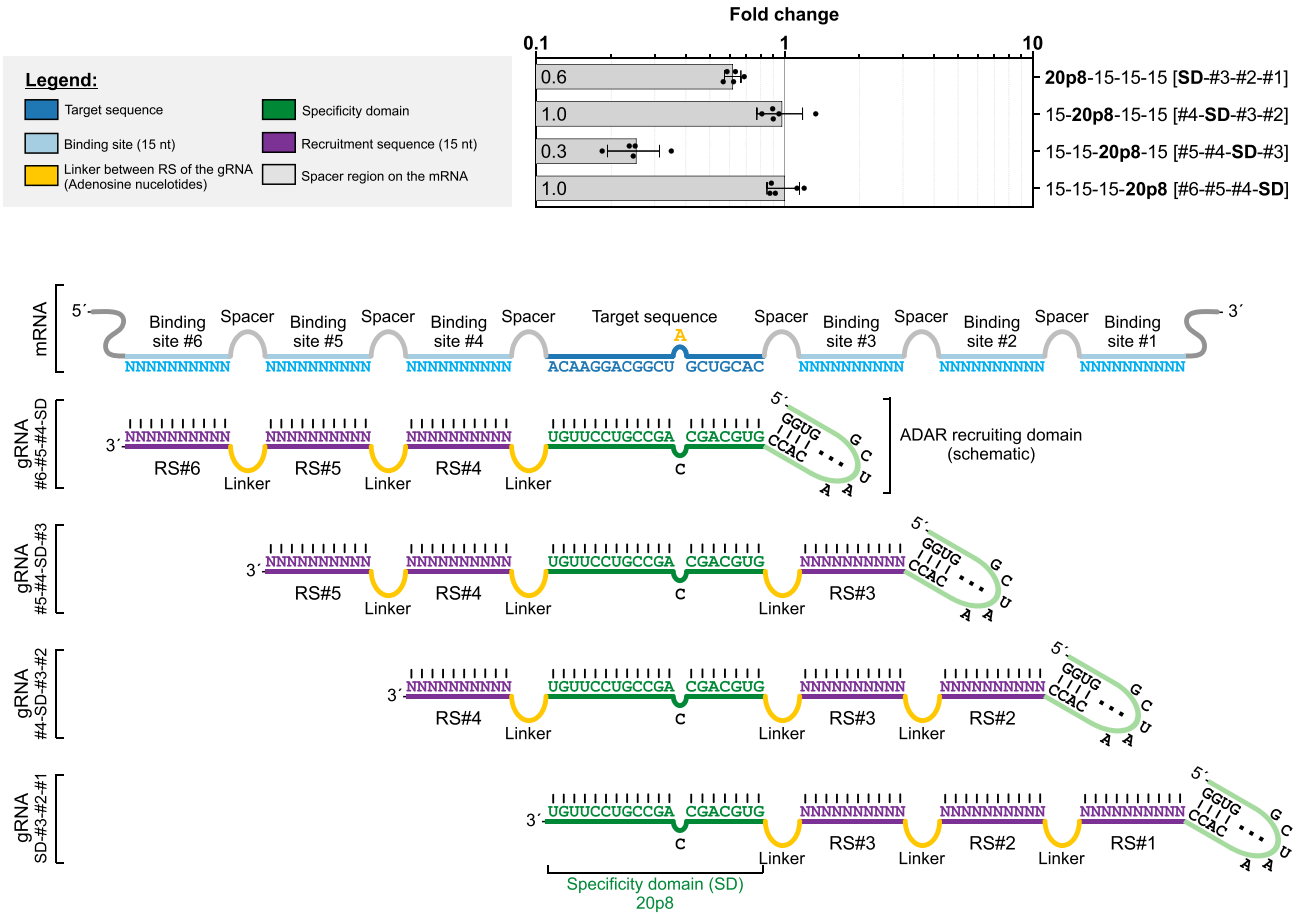


Extended Data Fig. 1 | Raw firefly- and renilla-luciferase RLU values of the dual-luciferase reporter system applied to different cell-lines via adenoviral transduction. The dual-luciferase reporter system is explained in Supplementary Fig. 2. The transduction settings are explained in the methods section. This figure contains the raw data from Fig. 1h. Data are shown as the mean ± s.d. of N = 5 biological replicates.



Extended Data Fig. 2 | Avoidance of misfolding of the guide RNA is important for editing. In this experiment the detrimental effects of strong secondary structures within the antisense part of a guide RNA were highlighted by comparing a set of CLUSTER guide RNAs in a dual-luciferase assay performed in HeLa cells. If a section of an antisense part within a gRNA is sequestered by another section of the same antisense part via backfolding, we call this 'masking'. Masking reduces the contact surface of the gRNA with the target mRNA. To highlight the problem and its effect on editing, a reference gRNA with 3xRS was compared to three further guide RNAs which contain additional sequences (masking RS, mRS) that step-wise increase the level of masking. For example, mRS#4 masks RS#3 by forming a perfect RNA duplex, inducing a strong secondary structure into the antisense part of the CLUSTER guide RNA. In the most extreme example (containing mRS#4-6), all three RS (#1-3) are masked leaving only the specificity domain for target binding. Accordingly, editing yields drop with increasing masking, highlighting that avoidance of secondary structure is important for the in-silico selection of CLUSTER guide RNAs. The respective folded guide RNA structures are shown, as predicted with the Nupack tool (J. N. Zadeh, et al. J Comput Chem 32, 170–173 (2011)), demonstrating that our in-silico approach generates CLUSTER guide RNA with a perfectly folded ADAR recruitment domain (R/G motif) but with very little secondary structure in the specificity domain and recruitment sequences. Data are shown as the mean ± s.d. of N = 5 biological replicates.

Extended Data Fig. 3 | The length of the adenosine linkers between recruitment sequences has only a moderate effect the editing yield. A set of 3xRS CLUSTER guide RNAs differing only in the length (0-10 nt) of the adenosine linkers between their recruitment sequences (RS) were compared in a dual-luciferase assay performed in HeLa cells. The data was normalized to the standard 3xA linker gRNA, which showed the highest editing yield. Data are shown as the mean ± s.d. of N = 5 biological replicates.

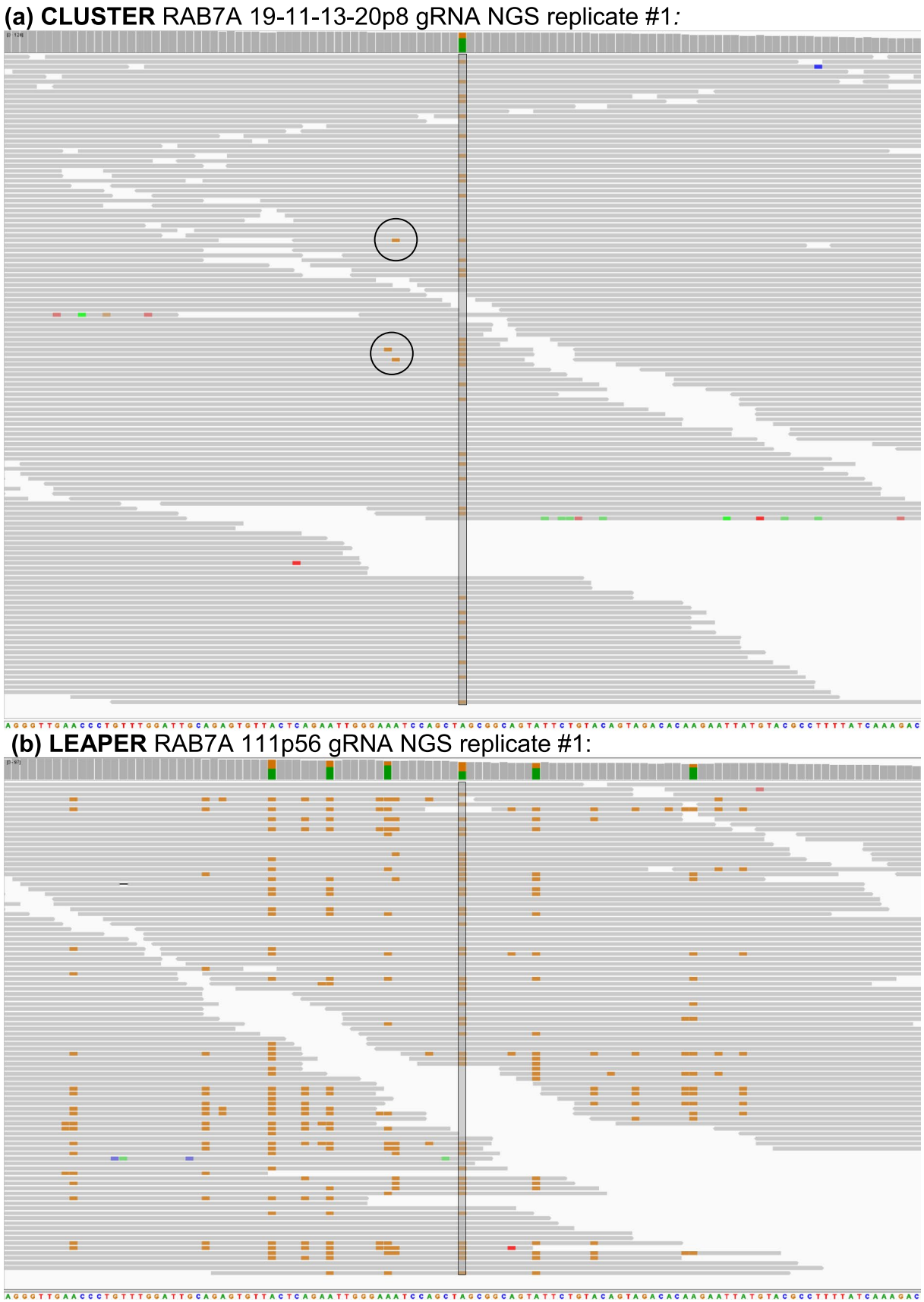


Extended Data Fig. 4 | The recruitment sequences (RS) and the specificity domain (SD) can be arranged with certain flexibility. The components of the antisense part of a CLUSTER guide RNA targeting the dual luciferase reporter in HeLa cells were newly arranged, starting from the conventional #6-#5-#4-SD design. Notably, some designs, for example #4-SD-#3-#2 gave comparably good editing yields, even though the SD is placed within the cluster of recruitment sequences. Importantly, the latter design allows to include the nucleotides space 3´ of the targeted adenosine (for example binding sites #3, #2, #1) for the binding of the CLUSTER guide RNA. Data are shown as the mean ± s.d. of N = 5 biological replicates.

Extended Data Fig. 5 | Benchmark of CLUSTER vs. LEAPER guide RNAs targeting the BMPR2 W298X amber mutation. A CLUSTER guide RNA containing nine recruitment sequences (RS#1-9) was compared to a 111 nt LEAPER guide RNA for targeting the said mutation delivered as a cDNA into HeLa cells. Although the on-target editing yield was similar, the LEAPER guide RNA showed widespread (≥14 sites) and severe (11-51% yield) bystander off-target events, while such off-target events are barely detectable for the CLUSTER guide RNA. Exemplary Sanger sequencing reads covering the complete binding sites of both guide RNAs are shown. Off-target events are highlighted by red arrows. The dataset in the benchmark table shows the mean ± s.d. of N = 3 biological replicates.

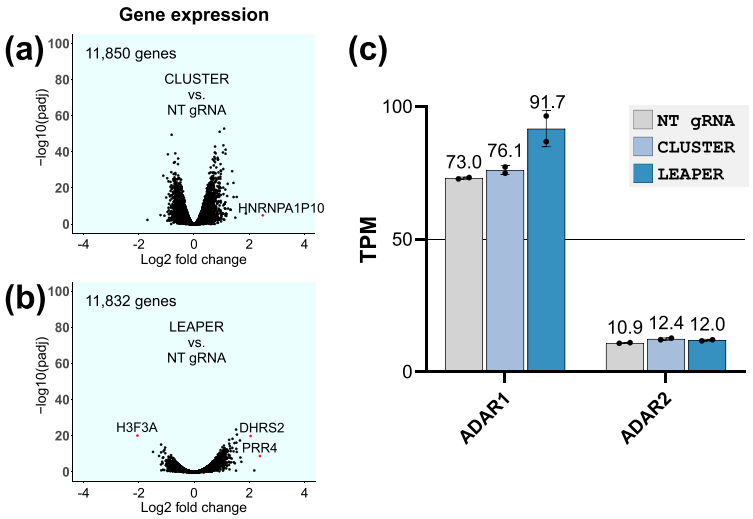
Gene	Position	Function	NT gRNA	CLUSTER	LEAPER
CTNNAL1	111,735.087	exonic	0.0%	12.2%	0.0%
HTATSF1	135,594,128	exonic	0.0%	10.6%	9.5%
RAB11FIP3	572,028	UTR3	27.1%	21.5%	48.5%
RAB7A	128,532,998	UTR3	0.4%	0.0%	17.2%
RAB7A	128,533,022	UTR3	0.5%	0.0%	35.8%
RAB7A	128,533,029	UTR3	1.3%	0.0%	23.0%
RAB7A	128,533,036	UTR3	1.3%	1.0%	21.5%
RAB7A	128,533,054	UTR3	0.4%	0.0%	29.7%
RAB7A	128,533,073	UTR3	0.9%	0.0%	17.8%
RAB7A	128,533,079	UTR3	0.9%	0.0%	11.8%
S100A10	151,966,272	UTR5	29.7%	17.5%	29.5%
WAC-AS1	28,810,277	ncRNA_exonic	30.1%	10.6%	12.0%
ZNF740	53,581,706	UTR3	0.0%	17.6%	0.0%

Extended Data Fig. 6 | Heat-map listing the off-target events induced by RNA editing with CLUSTER- and LEAPER-guide RNAs determined by next generation sequencing. This list shows the subset of off-target editing events at 'unknown' sites that have been detected by the pipeline searching for significantly differentially edited sites in at least one sample when cells were treated either with the CLUSTER or the LEAPER guide RNA and compared to the non-targeting control guide RNA (NT gRNA). Sites were assigned 'unknown' when they were not listed in the RADAR database. Nonsynonymous editing was only detected for a single site, HTATSF1 (S742G). Mapping analysis detected sites for potential off-target binding of the RAB7A guide RNAs to CTNNAL1, HTATSF1, and ZNF740. The detected RAB7A sites represent the fraction of bystander editing sites close to the on-target site that met the significance and cut-off criteria of the pipeline. Data are representing the mean ± s.d. of N = 2 NGS replicates.

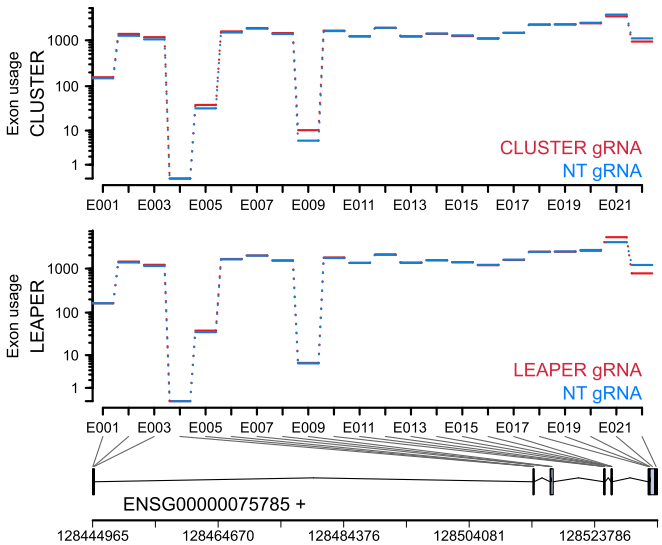


Extended Data Fig. 7 | See next page for caption.

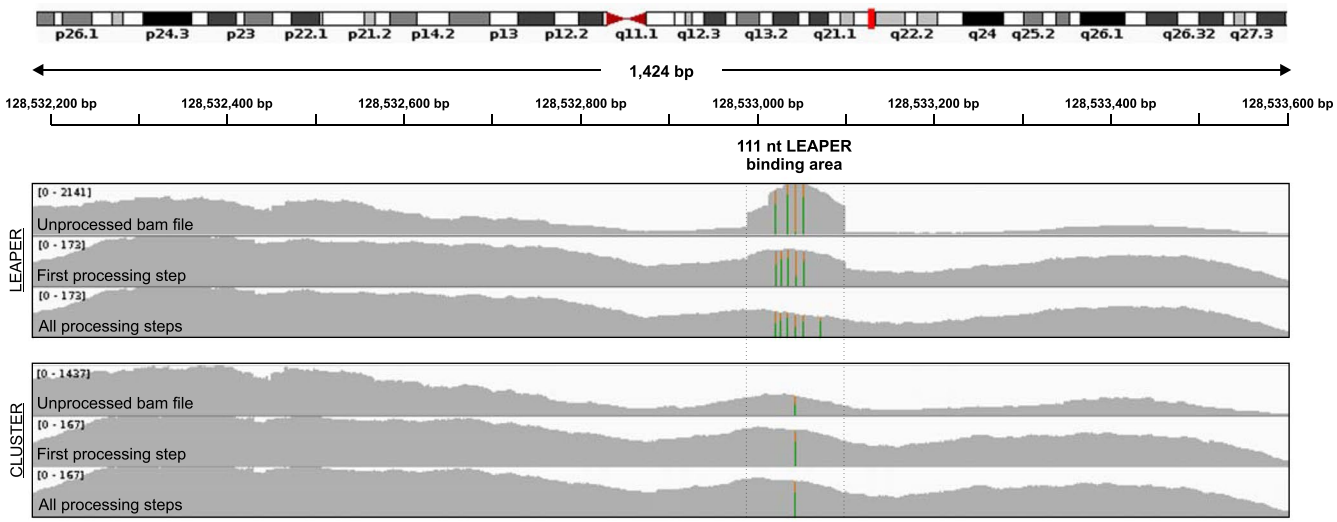
Extended Data Fig. 7 | Displaying major differences in bystander editing around the on-target site as determined by next generation sequencing. NGS reads mapped around the on-target site of RAB7A obtained from the editing with either a CLUSTER (**a**) or a LEAPER (**b**) guide RNA were displayed with the integrative genomics viewer (Robinson, J. T., et al. (2011). 'Integrative genomics viewer.' Nat Biotechnol 29(1): 24-26.). The analysis clearly shows the high precision of the CLUSTER approach and the massive bystander editing with the LEAPER guide RNA. On-target (centred bar) and Off-target A > G events are indicated by ochre (guanosine) squares. The very few detected bystander edits of the CLUSTER guide RNA are highlighted with two circles in panel (a). Red (thymidine), green (adenosine) and blue (cytidine) squares indicate single nucleotide mismatches not caused by RNA editing. The displayed area shows the full 111 nt binding region of the LEAPER gRNA that also includes the full 20 nt targeting sequence of the CLUSTER gRNA. Data derive from NGS replicate #1 for both the CLUSTER and the LEAPER guide RNA treatment.



Extended Data Fig. 8 | Gene expression changes in HEK293FT cells after treatment with CLUSTER- or LEAPER guide RNAs. (a) and **(b)** The volcano plots show the DESeq2 gene expression after CLUSTER or LEAPER guide RNA treatment relative to the non-targeting guide RNA (NT gRNA) treatment. Positive log2 fold change values indicate upregulated expression for the CLUSTER or LEAPER guide RNA treatment, respectively. Out of the ≈11.000 genes (TPM ≥ 2), only very few genes were detected as significantly differently expressed (padj threshold < 0.01 and|Log2 fold change | greater 2), as indicated by red dots. **(c)** The potential effect of guide RNA treatments (LEAPER versus CLUSTER versus non-targeting control guide RNA) on ADAR expression was estimated from the TPM values. No notable difference in ADAR expression between the three different treatments was detectable. We provide N = 2 data points in panel (c) representing the two respective NGS replicates, which have been analyzed independently.



Extended Data Fig. 9 | Editing did not affect the RAB7A mRNA exon usage. For both, the CLUSTER and the LEAPER guide RNA-treated samples of our NGS dataset, the exon usage of the RAB7A target mRNA was evaluated and compared to the samples treated with the non-targeting guide RNA control (NT gRNA). In both cases (CLUSTER and LEAPER), no significant difference (padj < 0.05 and |logfold| >1) in exon usage compared to the control was detectable. Each exon bin displayed in this figure shows the mean exon usage as determined by the DEXSeq method (Anders, S., et al. (2012). 'Detecting differential usage of exons from RNA-seq data.' Genome Res 22(10): 2008-2017) from two independent NGS replicates.



Extended Data Fig. 10 | NGS data processing to remove artefacts caused by the LEAPER treatment. For the LEAPER guide RNA NGS samples, we observed an abnormally high coverage around the RAB7A target region that corresponds to the binding area of the LEAPER guide RNA (111 nt). We found that such reads originated by the LEAPER guide RNA itself and were falsely aligned to this region, leading to this abnormal coverage pattern. To remove such reads, we applied a procedure (as described in detail in the Methods section) to the LEAPER and all other NGS samples. Here, the top tracks show the unprocessed bam files, visualized by the Integrative Genomics Viewer (IGV), for the LEAPER and CLUSTER guide RNA samples of NGS replicate #1, respectively. The middle tracks resulted from the first processing step that removed PCR duplicates in the 3'UTR of RAB7A with the rmdup command in samtools. The bottom tracks show the final bam files after all described processing steps were performed. These files were then used for the downstream analyses (see Methods section). When revisiting the NGS raw data from the original publication about LEAPER guide RNAs, we found the same phenomenon there (Qu et al., Nat Biotechnol, 2019).

nature research

Corresponding author(s): Thorsten Stafforst

Last updated by author(s): Sep 17, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No custom-made code was required to collect data. The Western blot pictures were taken using the Odyssey Fc imaging system (LI-COR). No further image processing was done with respect to brightness or contrast.
Data analysis	Mapping of RNA-seq and reads: STAR (version 2.5.3a) was used to align the reads to the hg19 reference genome. To remove clonal reads (PCR duplicates) mapped to the same location we ran Picard tools (version 1.129). Unique and non-duplicate reads were subjected to local realignment and base score recalibration using the IndelRealigner and TableRecalibration from the Genome Analysis Toolkit (GATK, version 3.6). Identification of editing sites from RNA-seq data. To remove Leaper guide RNA sequences that were falsely aligned to the targeting region, we used the rmdup command in samtools 1.11 to remove PCR duplicates in the RAB7A 3'UTR region. Additionally, we removed all reads that contained the sequence 'AAGGGTG' (3'end of the Leaper gRNA) and reads that ended with 'TCAAAGAC' (5'-end of the Leaper gRNA). As last step, we removed all reads that originated from the antisense sequence of the RAB7A gene. This procedure was applied to all samples (CLUSTER, LEAPER, ScrRNA). To call variants from the mapped RNA-seq reads, we used the UnifiedGenotyper from GATK. In contrast to the usual practice of variant calling, we identified the variants with relatively loose criteria by using the UnifiedGenotyper tool with options stand_call_conf 0, stand_emit_conf 0, and output mode EMIT_VARIANTS_ONLY. Variants from nonrepetitive and repetitive non-Alu regions were required to be supported by at least three reads containing mismatches between the reference genome sequences and RNA-seq. Supporting of one mismatch read was required for variants in Alu regions. This set of variant candidates was subject to several filtering steps to increase the accuracy of editing site calling. We first removed all known human SNPs present in dbSNP build 137 (except SNPs of molecular type "cDNA"; database version 135; http://www.ncbi.nlm.nih.gov/SNP/), the 1000 Genomes Project, and the University of Washington Exome Sequencing Project (http://evs.gs.washington.edu/EVS/). To remove false-positive RNA-seq variant calls due to technical artifacts, further filters were applied as previously described (see manuscript for fruther details). In brief, we required a variant call quality Q > 20, discarded variants if they occurred in the first 6 bases of a read, removed variants in simple repeats, removed intronic variants that were within 4 bp of splice junctions, and discarded variants in homopolymers. Moreover, we removed reads mapped to highly similar regions of the transcriptome by BLAT. Finally, variants were annotated using ANNOVAR based on gene models from Gencode, RefSeq, Ensembl and UCSC.

All sites identified from RNA-seq data were compared with all sites available in the RADAR database to be referred as 'known' sites if found in RADAR or 'novel' sites if not found.

Identification of significantly differently edited sites. We merged all sites found in the RNA seq samples with the sites of the RADAR database and quantified editing levels of edited sites with ≥ 50 reads coverage (combined coverage of both replicates) and performed Fisher's exact tests followed by Benjamini–Hochberg's multiple test correction (adjusted $P < 0.01$) to identify significantly differently edited sites across the samples (absolute editing difference > 10%).

Measuring RAB7A editing precision. To compare the specificity of CLUSTER and LEAPER guide RNA, we selected all mapped reads that contained the edited target sequence 'GCTGGCGG'. The edited reads and their partner reads were compared to the RAB7A sequence covering the editing region to identify A-to-G mismatches. As a control, we used the non-targeting sample to quantify A-to-G mismatches in the reads that covered the unedited target sequence 'GCTAGCGG'.

Gene expression analysis. We ran RSEM (version 1.2.21) on the STAR alignments to obtain read counts and TPM values to characterize gene expression. We used the read counts to perform expression analysis with DESeq for all genes with TPM ≥2 (for both replicates) and identified significantly expressed genes by using a threshold of padj<0.01 and an |Log2 fold change| greater than 2.

The DEXSeq exon usage analysis was performed as elucidated in "Anders, S., et al. (2012). "Detecting differential usage of exons from RNA-seq data." Genome Res 22(10): 2008-2017".

Non-NGS data were analyzed using Excel 2016 and GraphPad Prism 8, Figures were created with CorelDraw 2017, the manuscript was written with Word 2016, the custom "recruitment cluster finder" tool (https://github.com/recruitment-cluster-finder/rcf/releases) was written in python 3.7.2 and used the Vienna RNA package 2.0 (https://www.tbi.univie.ac.at/RNA/), gRNA folds were created using the nucleic acid package (Caltech, Nupack.org).
qPCR analysis was performed using the 7500 data analysis software v2.3.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Transcriptome-wide RNA-seq data are accessible via the NCBI GEO database with accession code GSE184244. The RADAR database (http://RNAedit.com) was used as reference to assess the editing homeostasis. The dbSNP database version 135 (http://www.ncbi.nlm.nih.gov/SNP/) was used to discriminate human SNPs from RNA editing events. NGS artefacts caused by the LEAPER gRNA treatment were confirmed within transcriptome-wide RNA-seq data from Qu et al. (NCBI Sequence Read Archive database, accession code PRJNA544353).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Experiments for evaluating editing yields via Sanger sequencing were mostly done in n=3 biological replicates in rare cases in n=2, n=4 or n=5. Data points are always displayed individually. The evaluation of the editing yields via the dual-luciferase assay for the animal experiment was done with n=3-5 mice per group, with 5 individually measured liver lobes per mouse and measured in three technical replicates per liver lobe. The evaluation of the editing yields or fold change via the dual-luciferase assay for cell culture experiments was done with n=5 or n=6 biological replicates with one technical measurement each. No sample size calculation was performed. The group sizes for cell culture and animal experiments were selected based on the prior knowledge of variation. NGS analysis was performed with two independent replicates per sample; the required sequencing depth was determined in two previous studies (Vogel et al. Nature Methods 2018, Merkle et al. Nature Biotechnology 2019) and saturated with 50 Mio 100 bp paired-end reads at ~30,000 detected transcripts. This sequencing depths was also similar to other recent papers on global off-target effects of site-directed RNA editing (Cox et al. Science 2017, Rosenthal et al. RNA Biol. 2018).
Data exclusions	No data was excluded.
Replication	All experiments could be reproduced, as shown in the manuscript, the number of replications and nature of replicates is always given in the figure caption.

Randomization	All samples were treated according to the same protocols side-by-side with the respective controls and thus, there was no requirement for randomization.
Blinding	no blinding was performed due to the involvement of several experimentators.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	ADAR1 antibody (α-ADAR1, mouse monoclonal IgG, Santa Cruz cat. no.: sc-73408, clone no. 15.8.6, lot no. C2514, used in 1:1000 dilution) against amino acids 440-826 corresponding to the middle region of ADAR1 of human origin, Clone AC-15 (α-ACTB, mouse monoclonal IgG, Sigma Aldrich cat. No.: A5441) against Actin N-terminal peptide, Ac-Asp-Asp-Asp-Ile-Ala-Ala-Leu-Val-Ile-Asp-Asn-Gly-Ser-Gly-Lys. Anti-mouse–HRP antibody (Peroxidase AffiniPure Goat Anti-Mouse IgG (H+L), Jackson Immuno Research Laboratories, #115-035-003), Mouse Anti-Adenovirus Type 5 E1A (Clone M58 (RUO), BD Pharmingen, # 554155).
Validation	ADAR1 antibody: Validated in our lab via siRNA KO and Western Blot (in several cell lines), and by overexpression / Western blot PMID: # 28669490 PMID: # 28278381 PMID: # 27573237 PMID: # 27907896. Clone AC-15 antibody: PMID: # 15809369 PMID: # 15048076 PMID: # 21217779. Clone M58 (RUO) antibody: Validated in our lab via Western Blot (in 293 cells), PMID: # 3028247, PMID: # 3894685, PMID: # 3317834, PMID: # 2521301. Peroxidase AffiniPure Goat Anti-Mouse IgG (H+L) antibody: Validated in our lab via Western Blot (in several cell lines), Two examples of the 1181 citations registered on September 9th 2021: PMID: # 32319599, PMID: # 32782498.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HeLa (cat. no. CCL-2, ATCC) A549 (cat. no. 86012804, European Collection of Authenticated Cell Cultures ECACC) HepG2 (cat. no. ACC180, DSMZ, Braunschweig, Germany) Huh7 (cat. no. 300156, CLS GmbH, Heidelberg) SH-SY5Y (cat. no. CRL-2266, ATCC) SK-N-BE(2) (cat. no. CRL-2271, ATCC) U87MG (cat. no. HTB-14, ATCC) HEK-Flp-In T-Rex-A1p110 (cat. no. R78007, Thermo Fisher scientific, stably transfected with ADAR1 p110 vector in our lab) HEK-Flp-In T-Rex-A1p150 (cat. no. R78007, Thermo Fisher scientific, stably transfected with ADAR1 p150 vector in our lab) HEK-Flp-In T-Rex-ADAR2 (cat. no. R78007, Thermo Fisher scientific, stably transfected with ADAR2 vector in our lab) 293FT (cat. no. R70007, Thermo Fisher scientific)
Authentication	Authentication via STR profiling was performed by the commercial suppliers before purchase of the material. Cell lines were not additionally authenticated by us.
Mycoplasma contamination	HeLa, A549, Huh7, SH-SY5Y and SK-N-BE cell lines have been tested as mycoplasma-free by the commercial suppliers and in house. The HepG2 and U87MG cell lines have been tested as mycoplasma-free by the commercial suppliers, but were not again tested in house.
Commonly misidentified lines (See ICLAC register)	none were used.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Species: mouse, strain: C57BL/6. The negative, positive and editing group 2 (20-15-15-20p8 gRNA treatment) consisted in equal shares of male and female animals.

	The editing group 1 (15-13-11-20p8 gRNA treatment) consisted only of female animals. The animals were 5-6 weeks old.
Wild animals	No wild animals were used in this study.
Field-collected samples	No field-collected samples were used in this study.
Ethics oversight	The mouse experiments complied with animal experimentation regulations of Germany and were approved by the Ethics Review Committee of the regional council in Tübingen (Reference number 35/9185.81-2 / M 2/18)

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Supporting Information of Manuscript 3

CLUSTER guideRNAs enable precise and efficient RNA editing with endogenous ADAR enzymes in vivo

Philipp Reautschnig, Nicolai Wahn, Jacqueline Wettengel, Annika E. Schulz, **Ngadhnjim Latifi**, Paul Vogel, Tae-Won Kang, Laura S. Pfeiffer, Christine Zarges, Ulrike Naumann, Lars Zender and Thorsten Stafforst, *Nature Biotechnology*, 2022, *40.5*, 759-768

nature portfolio

https://doi.org/10.1038/s41587-021-01105-0

Supplementary information

CLUSTER guide RNAs enable precise and efficient RNA editing with endogenous ADAR enzymes in vivo

In the format provided by the authors and unedited

Supplementary Information

CLUSTER guide RNAs enable precise and efficient RNA editing with endogenous ADAR enzymes in vivo

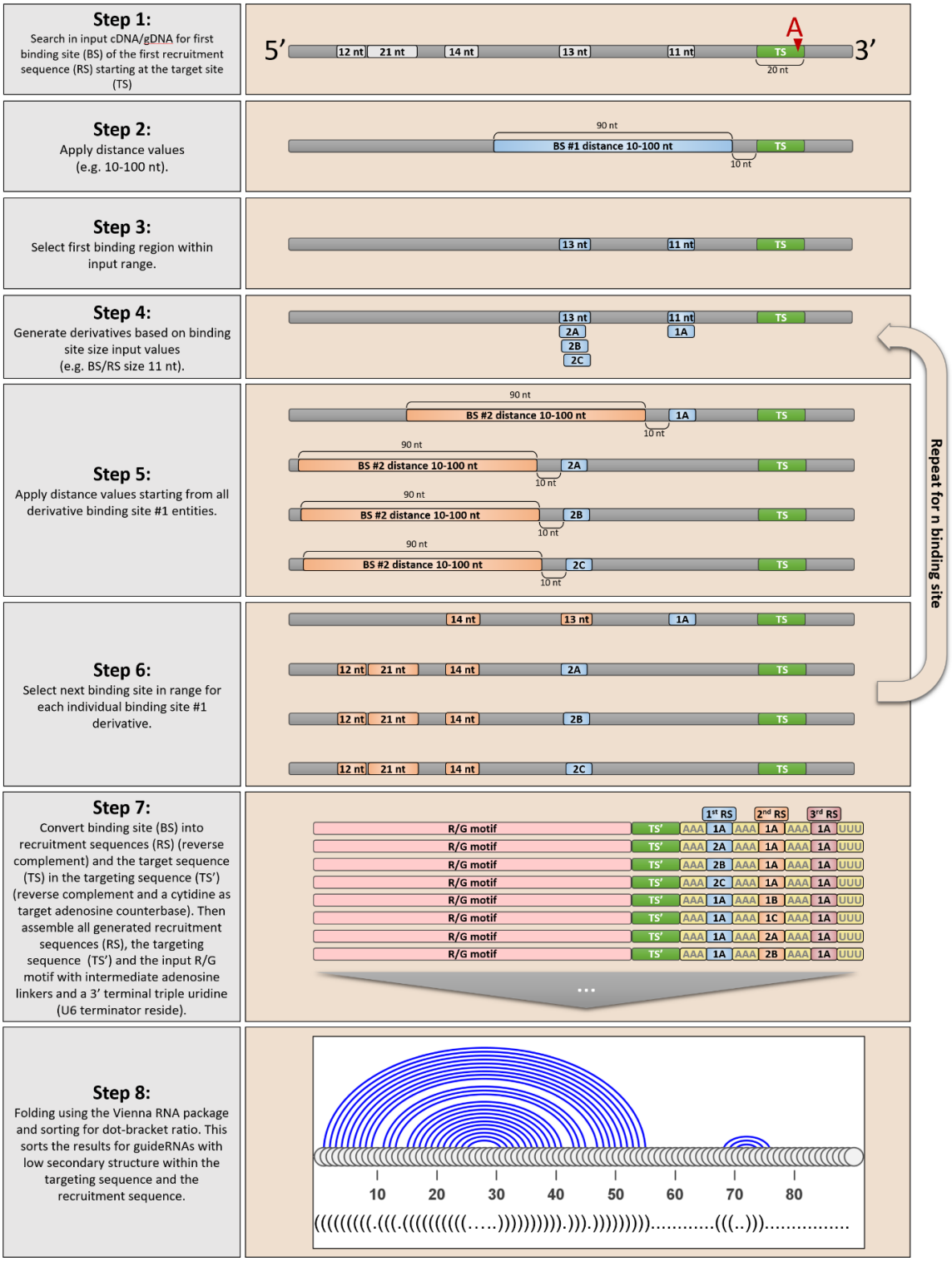
Philipp Reautschnig,[1] Nicolai Wahn, Jacqueline Wettengel,[1] Annika E. Schulz,[1] Ngadhnjim Latifi,[1] Paul Vogel,[6] Tae-Won Kang,[2,3] Laura S. Pfeiffer, [1] Christine Zarges,[1] Ulrike Naumann,[5] Lars Zender[2,3,4], Jin Billy Li[6] and Thorsten Stafforst[1]

Correspondence should be addressed to T.S
(thorsten.stafforst@uni-tuebingen.de)

Inventory of Supplementary Information

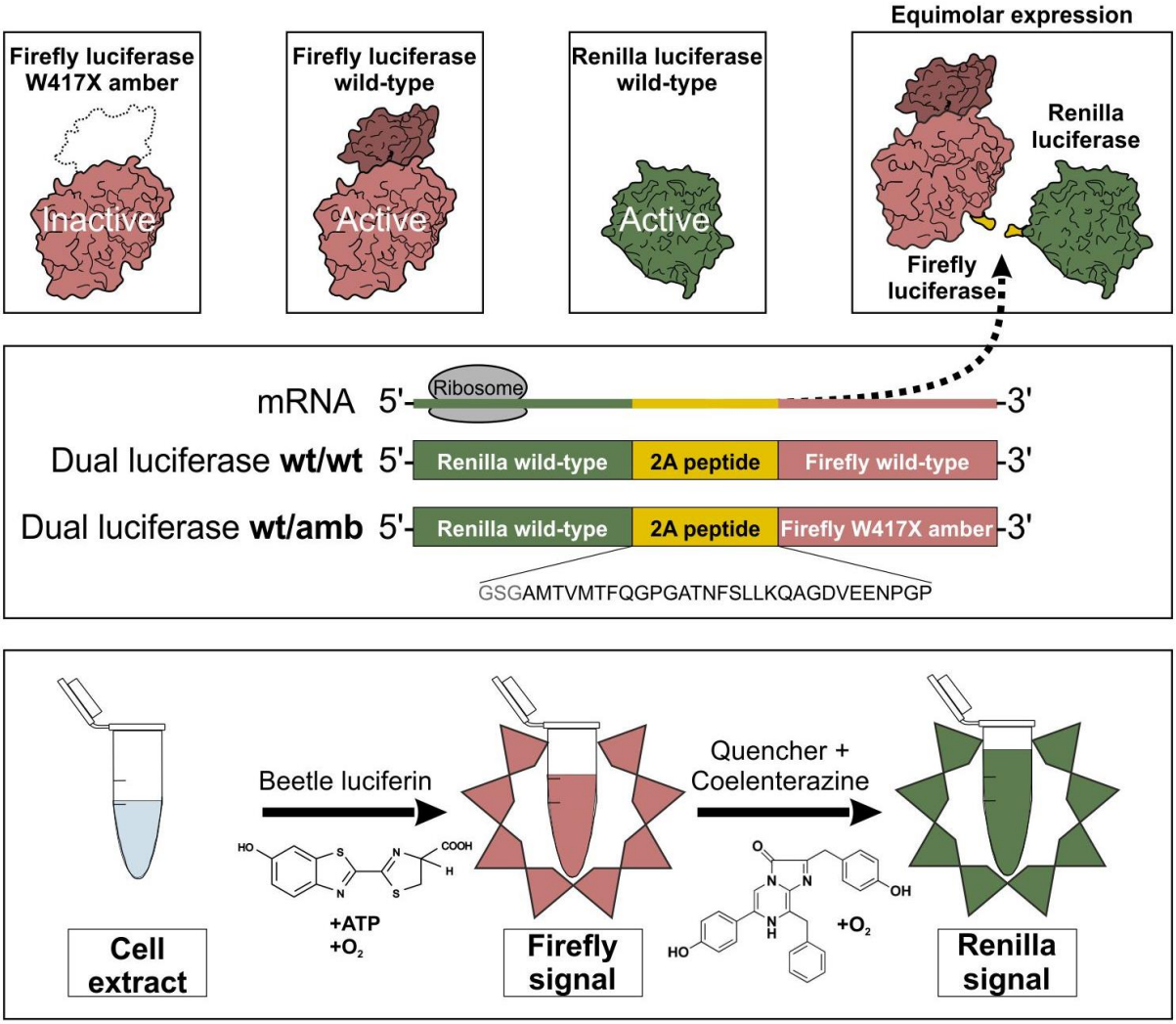
- Supplementary Figures S1-S16
- Supplementary Tables 1-5
- Supplementary Notes 1-2

Recruitment sequence *in-silico* optimization using the recruitment cluster finder tool



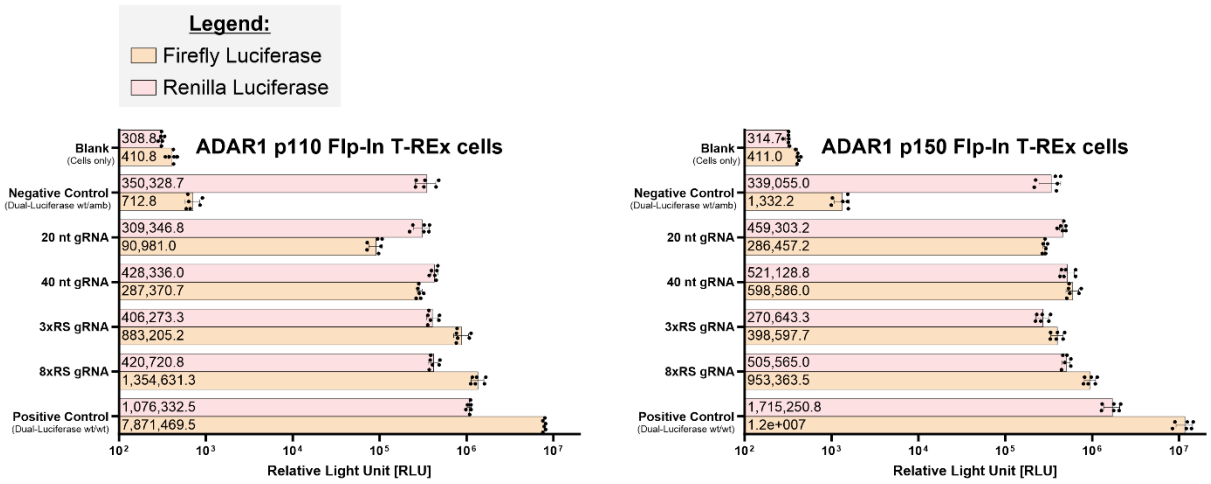
Supplementary Figure S1: Conceptual description explaining how the recruitment cluster finder (RCF) predicts CLUSTER guide RNAs. *This figure does not describe the exact combinatorial implementation, which is used by the algorithm to process input data, but explains the idea behind the tool in a conceptual way.*

Supplementary Figure S1 continued: Conceptual description explaining how the recruitment cluster finder (RCF) predicts CLUSTER guide RNAs. (Step 1) *The gDNA/cDNA corresponding to the target mRNA is screened for binding sites that contain only G, C, T and 5´GA Each binding site is defined by base distance indices relative to the beginning of the cDNA/gRNA input sequences. In the current version (1.0.1), the screening is performed 5' of the target site.* ***(Step 2)*** *Starting from the target site, binding sites in 5' range are detected.* ***(Step 3)*** *These binding sites are selected and analysed for their size.* ***(Step 4)*** *The binding site size input is used to generate binding site derivatives. If an uninterrupted binding site is, e.g., 13 nt long, then three derivative binding site can be created if the input binding site size was set to 11 nt.* ***(Step 5)*** *Starting from the first set of derivative binding sites, the next binding sites in range are detected.* ***(Step 6)*** *These binding sites are selected and analysed for their size. Steps 4-6 are repeated until n binding sites are selected.* ***(Step 7)*** *The resulting list of binding sites, which are matching all input variables are recombined and assembled with the input ADAR recruiting domain (e.g. R/G motif), adenosine linkers and the three terminal uridines, which result from the U6 termination signal.* ***(Step 8)*** *The Vienna RNA package is used to fold all guide RNAs within the list and to generate dot-bracket representations of these folds. The RCF allows to sort the structures by their free energy or by their dot-bracket ratio (ratio of the dot-bracket notation). Structures with good dot-bracket ratios (minimal base pairing within the antisense part of the guide RNA), are further sorted for the shortest number of brackets in a row within the antisense part. The shortest ones, which represent the weakest secondary structures, get the highest ranking. The selected recruitment sequences of resulting gRNAs with good secondary structures can then be manually BLASTed to exclude sequences which might cause off-target effects within the transcriptome.*

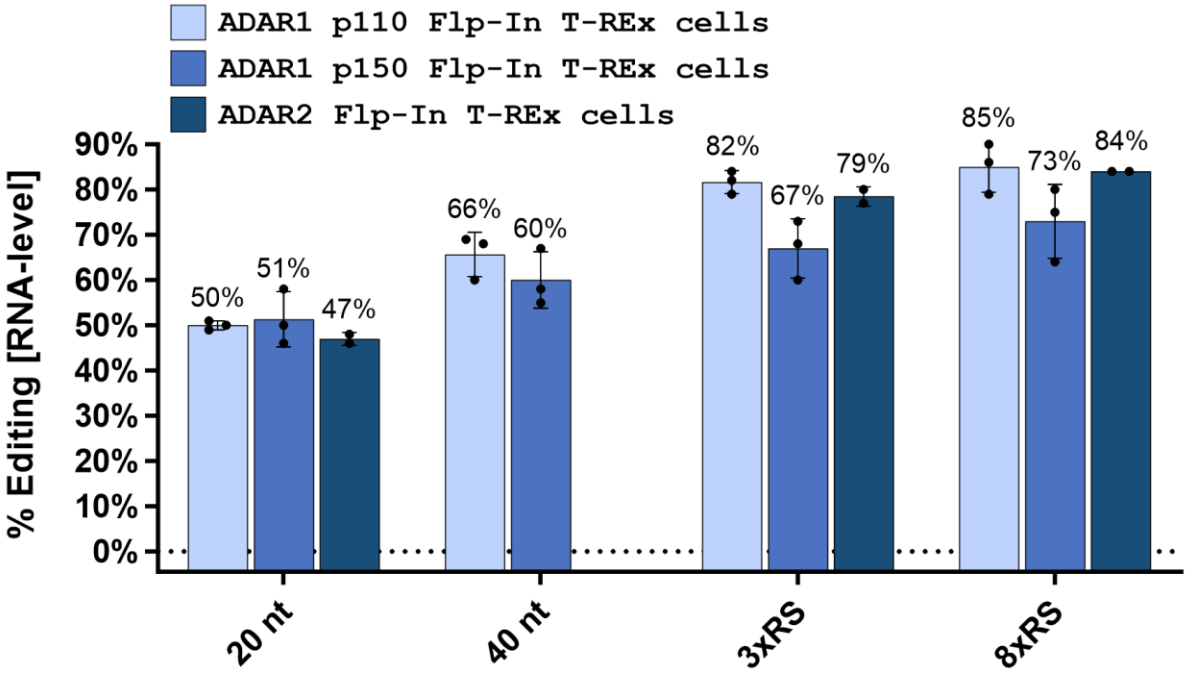


Supplementary Figure S2: The dual-luciferase reporter assay. *The 2A peptide (see J.H. Kim et al., PLoS One 2011, 6(4): e18556) allows the equimolar expression of both luciferases (firefly and renilla) from one bicistronic construct. This allows to eliminate the transfection bias by normalizing the firefly luminescence to the renilla luminescence. The one-pot dual luciferase assay is performed in a plate-reader by automated, successive addition of the substrates, followed by luminescence measurement. First the firefly substrate is added, followed by a firefly quencher and the renilla substrate. The reporter construct (wt/amb) combines a wildtype renilla with a firefly luciferase containing an editable premature stop codon. After editing this stop codon, firefly luciferase activity is switched on. By defining the renilla-normalized firefly luminescence from the wt/amb reporter as 0% and the respective luciferase signal from the wt/wt positive control construct as 100%, the percentage of restored firefly luciferase activity can be calculated. Furthermore, fold-changes between different guide RNA treatments can be calculated by setting the renilla-normalized firefly luminescence from one treatment in ratio to that from another treatment.*

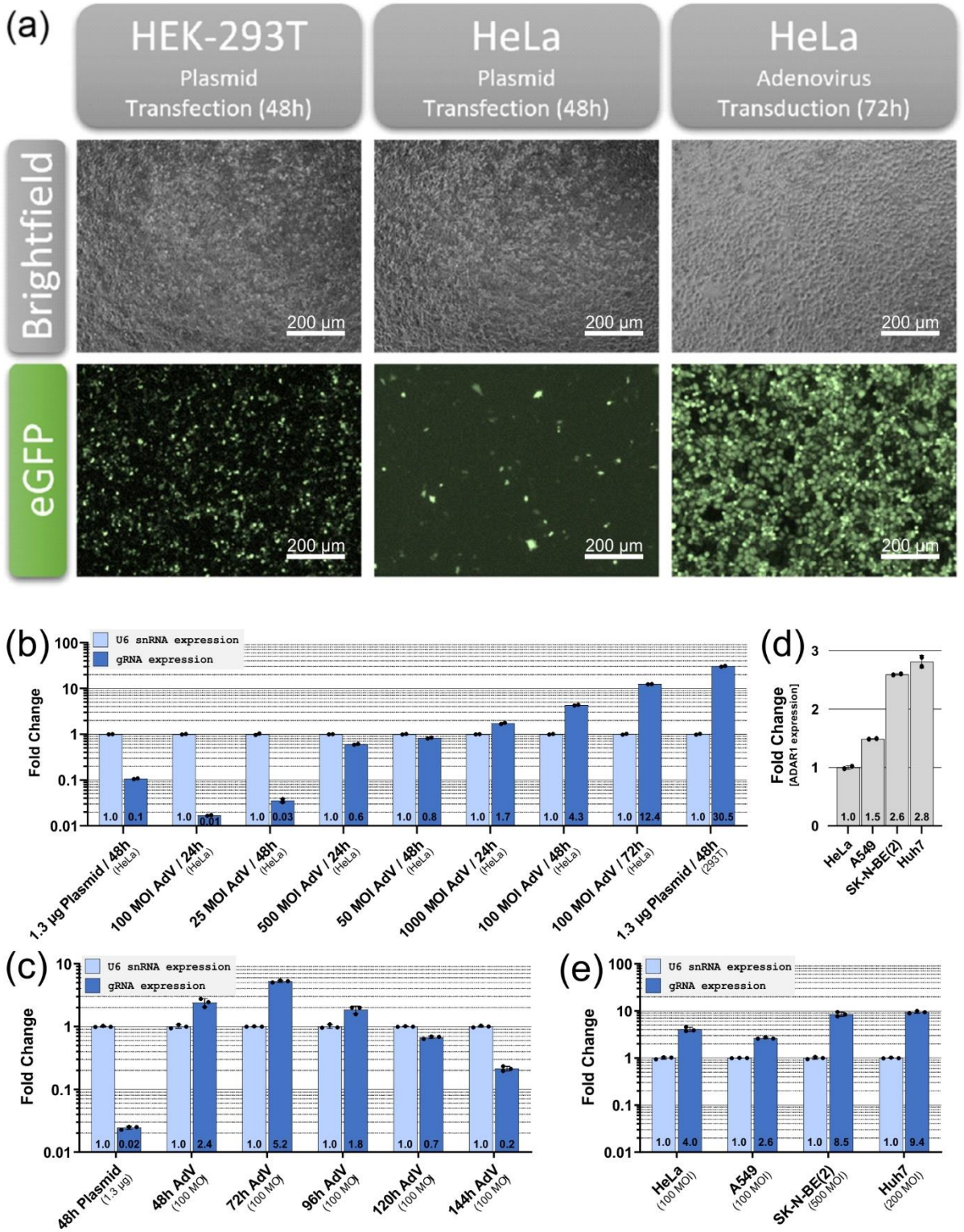
Supplementary Figure S3: Characterization of doxycycline-inducible transgene expression in engineered ADAR1 Flp-In 293 T-REx cells. *The cells were induced for 96h with 10 ng/ml doxycycline in DMEM+10%FBS. Medium was changed daily. 10 µg protein-lysate were loaded per lane. The anti-ADAR1 antibody was used in a 1:1000 dilution. The anti-β-actin antibody was used in a 1:5000 dilution. The image was taken with 20 seconds exposure time. No further image processing with respect to contrast or brightness was done. Non-induced cells were used as negative controls. Parental Flp-In 293 T-REx cells were used as reference for the endogenous ADAR1p110 expression levels. ADAR1 p110 and ADAR1p150 Flp-In T-REx cells seem to express their respective transgene at similar levels, under tight control of doxycycline. Expression of ADAR1p150 did not increase expression of the p110 isoform beyond its endogenous expression. N = 2 experiments were performed with similar results.*



Supplementary Figure S4: Raw firefly- and renilla-luciferase RLU values of the dual-luciferase reporter system applied to ADAR1 Flp-In T-REx cell-lines via plasmid transfection. *The dual-luciferase reporter system is explained in Fig. S2. The transduction settings are explained in the methods section. This figure displays the raw data from Fig. 1b and includes further controls, e.g. the Negative Control, where the reporter plasmid was transfected without a guide RNA plasmid. Data are shown as the mean ± s.d. of N = 6 biological replicates.*

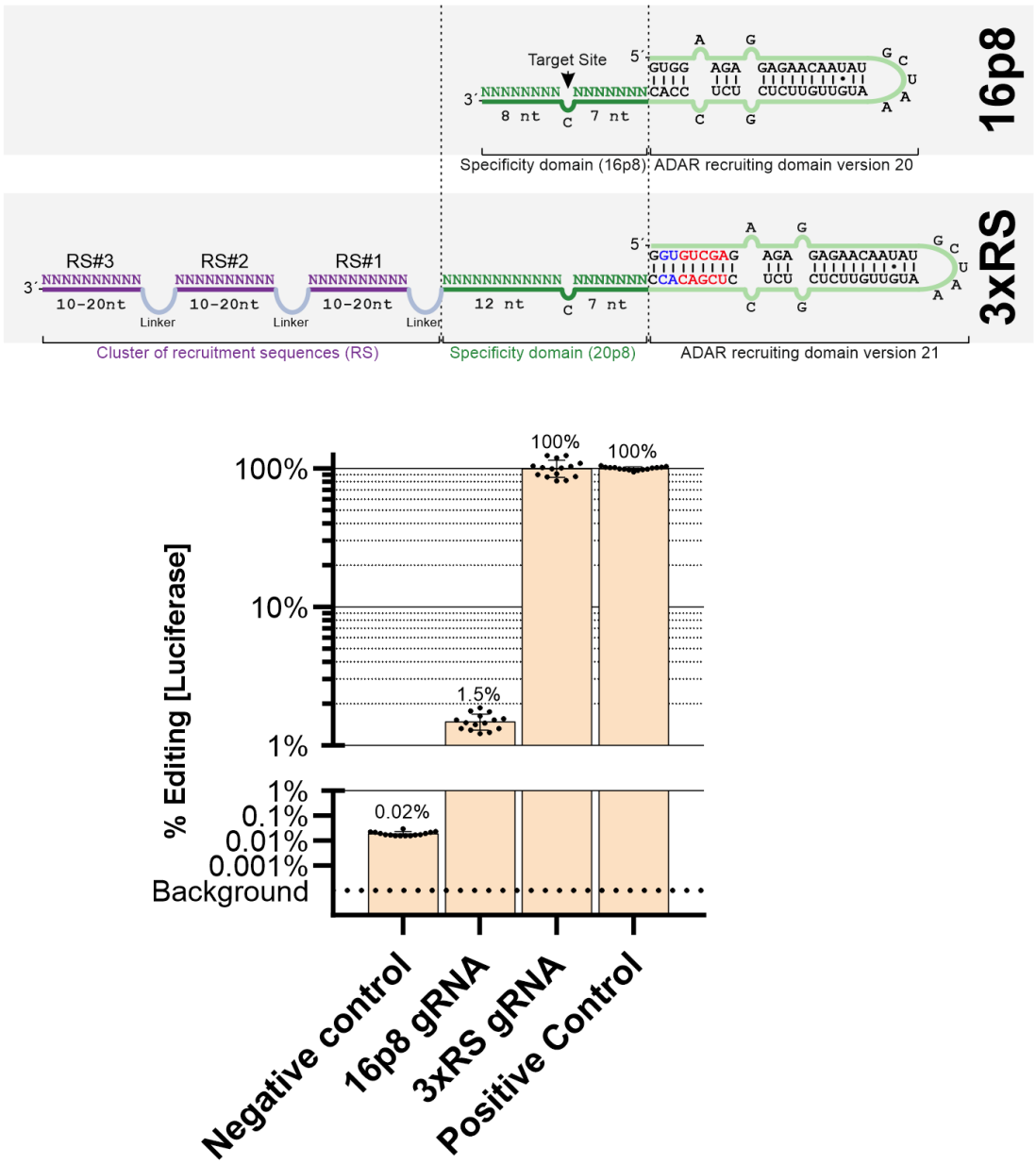


Supplementary Figure S5: Performance of in-silico optimized recruitment sequences for site-directed RNA-editing with overexpressed ADAR isoforms. *Editing of the dual-luciferase reporter in Flp-In T-REx cells overexpressing either ADAR1 p110, ADAR1 p150 or ADAR2. Prior design guide RNAs with 20 nt and 40 nt specificity domains were compared to 3x or 8x RS-containing gRNAs. This figure is an extension of main text Fig. 1c. Data are shown as the mean ± s.d. of N = 3 biological replicates.*

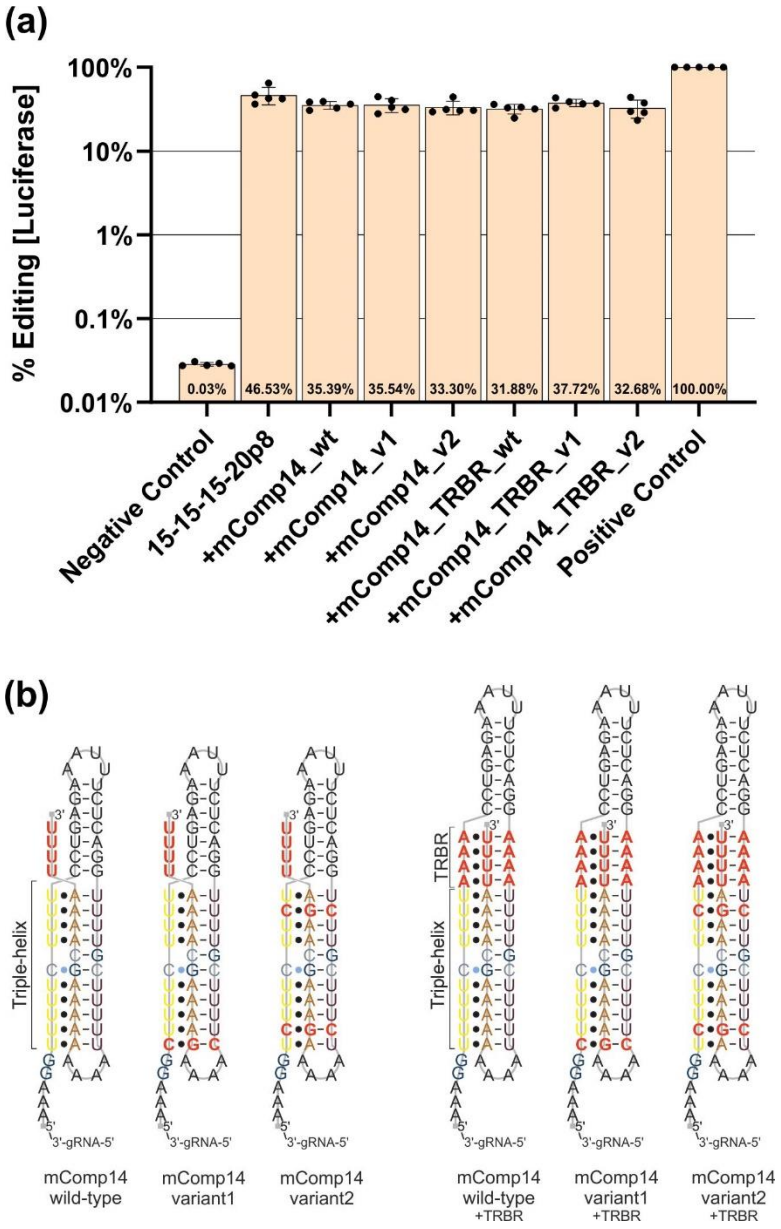


Supplementary Figure S6: Comparison of plasmid versus adenoviral vectors for guide RNA delivery in 293T and HeLa cells via microscopy and qPCR. (a) *Microscopy of eGFP expression after plasmid transfection versus adenovirus transduction. The plasmid-transfected HEK-293T cells were treated with 1300 ng gRNA and 300 ng eGFP using Lipofectamine 2000 at a 1:3 ratio (3 µl Lipofectamine 2000 per 1 µg of plasmid). The plasmid-transfected HeLa cells were treated with 1300 ng gRNA and 300 ng eGFP using Lipofectamine 3000 at a 1:1.5 ratio (1.5 µl Lipofectamine 3000 per 1 µg of plasmid).*

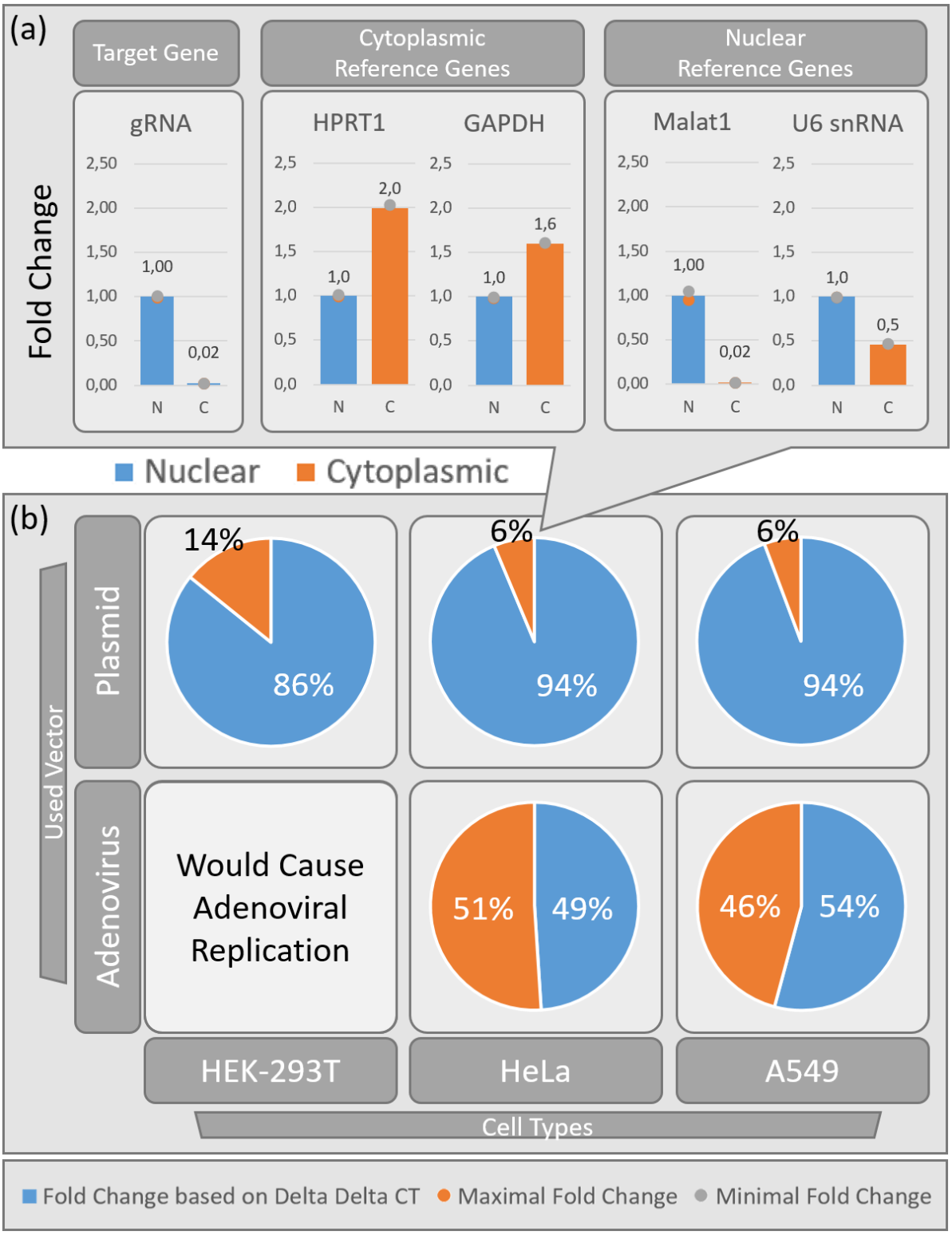
Supplementary Figure S6 continued: Comparison of plasmid versus adenoviral vectors for guide RNA delivery in 293T and HeLa cells via microscopy and qPCR. *The transduced HeLa cells were treated with 100 MOI of adenovirus, that encoded a β-actin 3' UTR targeting gRNA without RS and an eGFP-tag. All cells were seeded to be 70-90% confluent at the time of transfection/transduction, using 24-well scale cell culture plates. Images were taken at time points (48h or 72h) of highest transgene expression under the respective setting.* ***(b)*** *Comparison of guide RNA expression via qPCR depending on cell line, delivery vehicle, incubations times, and applied MOI. The transfections of 293T and HeLa cells with the identical plasmid-encoded gRNA served as reference. The transduced 16p8 guide RNA (16 nt long specificity domain, editing position 8, R/G motif version 1) was targeting the β-actin 3' UTR. The endogenous U6 snRNA was used as reference gene for normalization.* ***(c)*** *As b), but exploring the time-dependent expression profile for an adenovirally transduced (100 MOI) guide RNA in HeLa cells.* ***(d)*** *Relative endogenous ADAR1 expression between four different cell lines after transduction with 100 MOI (HeLa, A549), 200 MOI (Huh7) or 500 MOI (SK-N-BE(2)) of gRNA AdV.* ***(e)*** *As b), but exploring the relative guide RNA expression in different cell lines transduced with different MOI adenovirus with readout after 72h. The eGFP microscopy in a) is an example from a single experiment, but N = 3 experiments have been performed with similar results. The qPCR experiments in panel b) and d) are shown as the mean ± s.d. of N = 2 technical replicates. The qPCR experiments in panel c) and e) are shown as the mean ± s.d. of N = 3 technical replicates. The gRNA expression is normalized to U6 snRNA and given as a fold change relative to endogenous U6 snRNA. The ADAR1 expression is given as a fold change relative to the expression level in HeLa cells and normalized to β-actin.*



Supplementary Figure S7: Comparison of the prior design guide RNA (16p8) with the novel 3xRS CLUSTER design for harnessing endogenous ADAR in HeLa cells. *The upper part of this figure shows the design of the guide RNAs compared. The prior art design 16p8 guide RNA (Wettengel et al., Nucl. Acids Res. 45, 2797-2808 (2017), M. Heep, et al. Genes 8, 34 (2017)) contained a short (16 nt) specificity domain and an ADAR recruiting domain version 20. In comparison, the novel 3xRS CLUSTER design (15-13-11-20p8) consisted of a 20 nt long specificity domain, extended by three recruitment sequences (RS). The ADAR recruiting domain version 21 is extended by 5 bp (red) and two basepairs have been swapped against each other (blue) compared to version 20. The lower part of the figure shows the editing yields, determined using the dual-luciferase assay. 2x10^5 HeLa cells were seeded in 96 well scale and either reverse transduced with 175 MOI gRNA (16p8 or 3xRS gRNA AdV) or not treated until the next day. After 24h, the negative control and editing wells were transduced with 175 MOI dual-luciferase wt/amb adenovirus. The positive control wells were treated with 5 MOI dual-luciferase wt/wt adenovirus. 96h post seeding the dual-luciferase assay was performed. Data are shown as the mean ± s.d. of N = 5 biological replicates measured in technical triplicate. The firefly luminescence was normalized to renilla luminescence and set in ratio to the positive control*

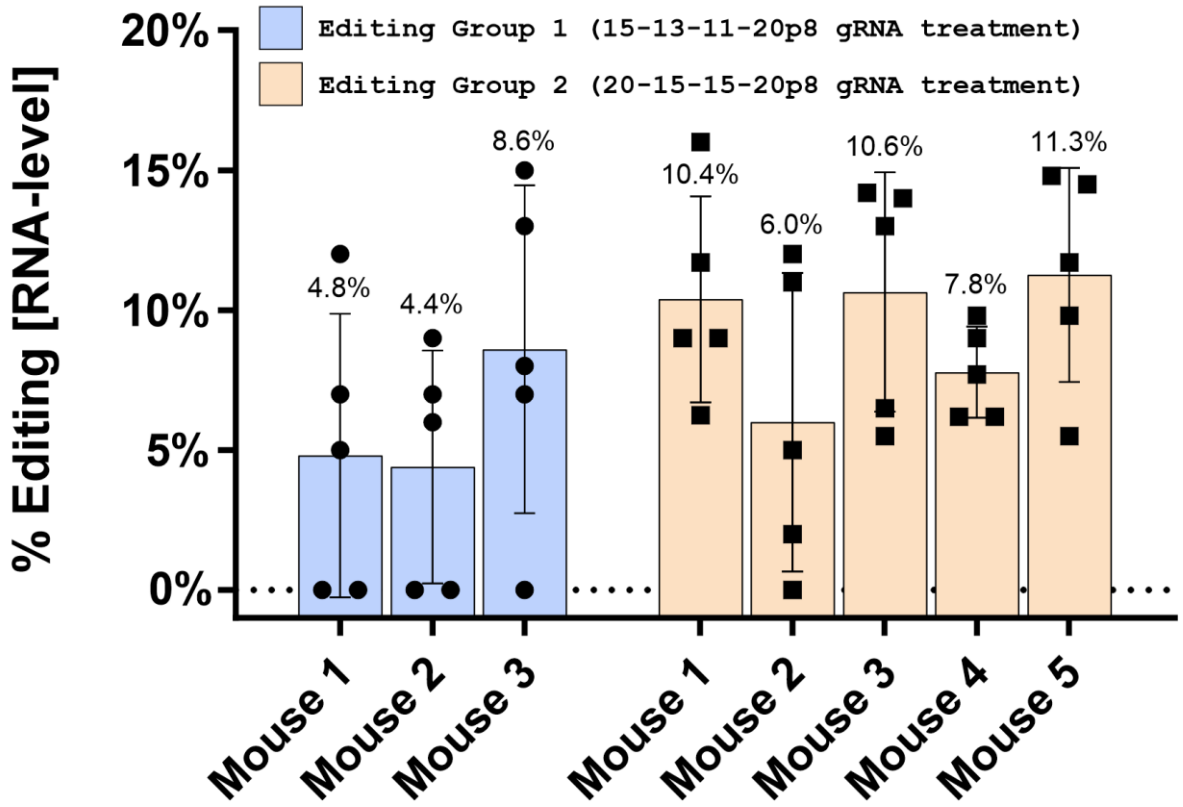


Supplementary Figure S8: Additional 3' terminal triple-helix forming elements do not improve the performance of CLUSTER guide RNAs. (a) *CLUSTER guide RNAs (15-15-15-20p8 design) targeting the dual-luciferase reporter in HeLa cells were – or were not – modified with additional 3´-terminal motifs and analyzed for in the reporter assay 6 days post transfection. Plotted are the % restored normalized luciferase activity in HeLa cells. Six different murine Comp14 triple-helix motifs were tested, for structures see b). None of them gave better editing performance.* ***(b)*** *Given are the structures of all six tested motifs. The structures are based on a reported motif (mComp14 wildtype) and were varied to remove the U6 terminator motif and to sequester in the 3´terminus inside the helix in what we call a terminator residue binding region (TRBR). Conrad, N. K. (2014). "The emerging role of triple helices in RNA biology." Wiley Interdiscip Rev RNA 5(1): 15-29. Data are shown as the mean ± s.d. of N = 5 biological replicates.*

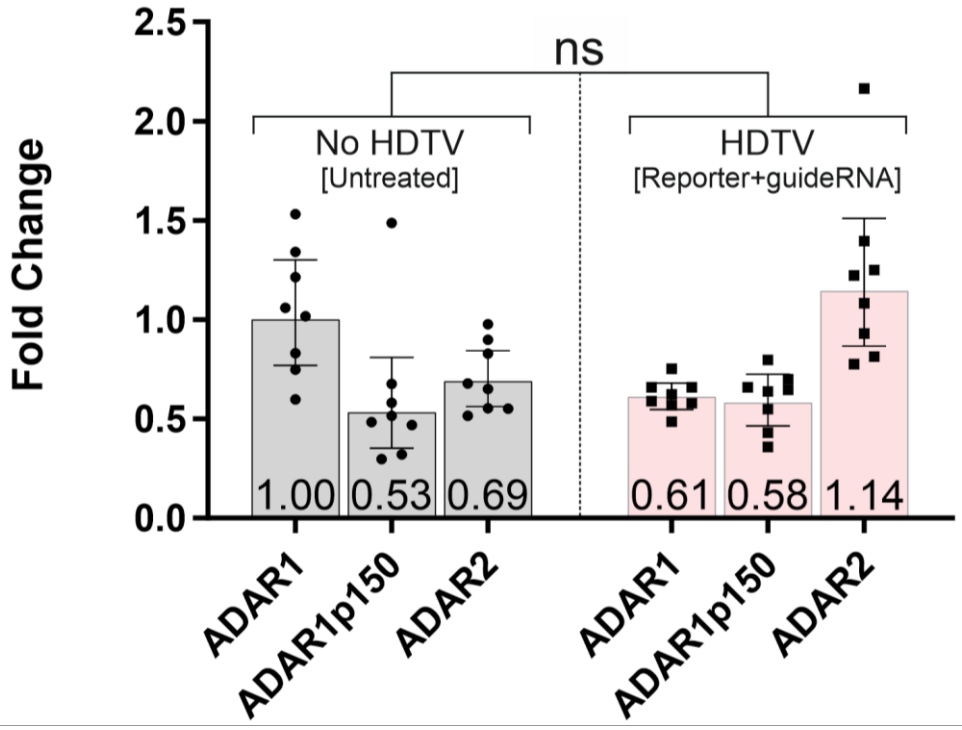


Supplementary Figure S9: Sub-cellular localization of guide RNAs depending on cell line and delivery vehicle. *For this experiment, cyto- and nucleoplasm was fractionated after guide RNA delivery, and the relative amount of guide RNA in cyto- and nucleoplasm was determined by qPCR. Fractionation was controlled with qPCR of housekeeping genes (MALAT1 & U6 nuclear; GAPDH & HPRT1 cytosolic). The 16p8 guide RNA (16 nt long specificity domain, editing position 8, R/G motif version 1) targeting the β-actin 3' UTR was either transfected as plasmid or transduced via an adenovirus.*

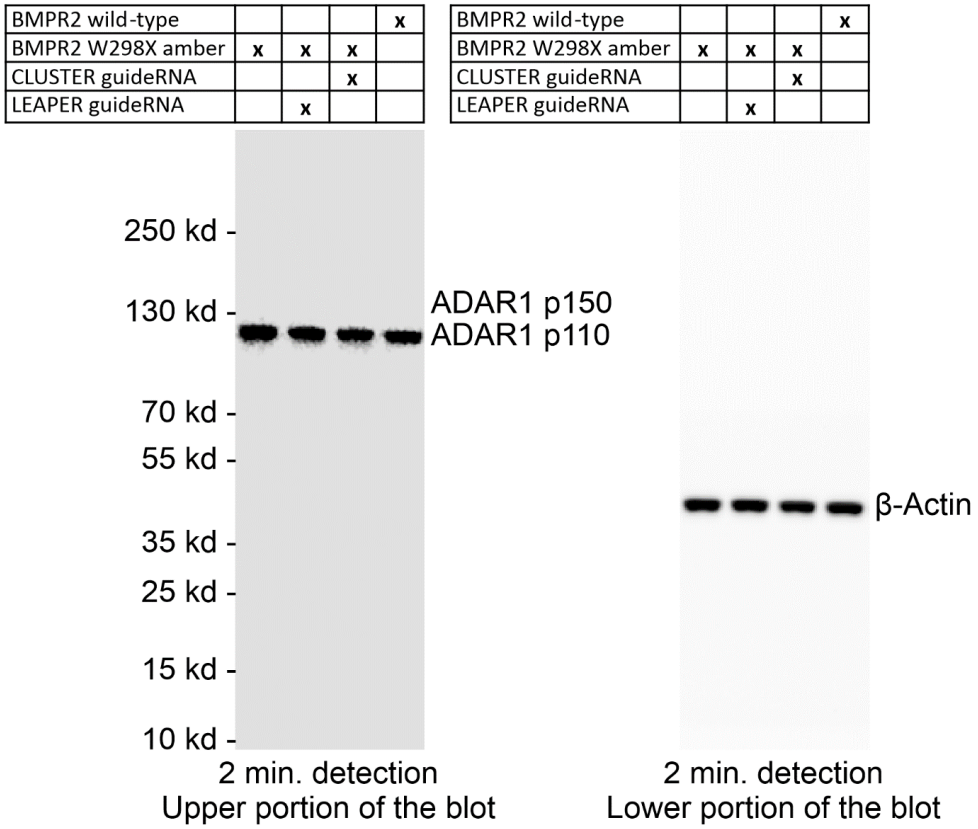
Supplementary Figure S9 continued: Sub-cellular localization of guide RNAs depending on cell line and delivery vehicle. *The geometric mean of all four reference genes was used to normalize the results for the guide RNA.* ***(a)*** *Exemplary fold-change results for the target gene (guide RNA) and the reference genes after fractionation.* ***(b)*** *Nucleo- versus cytosolic distribution of the guide RNA depending on the used vector and cell type. The plasmid-borne delivery gave largely nuclear expression of the guide RNA for all three cell lines. In contrast, adenoviral delivery gave considerably more guide RNA expression in the cytosol. $2x10^5$ HEK-293T cells per well were seeded in 24-well scale. 24h post seeding, the cells were transfected with 1300 ng gRNA plasmid using a Lipofectamine 2000 ratio of 1:3. $1x10^5$ A549 cells, or $6x10^4$ HeLa cells, per well were seeded in 24-well scale. 24h post seeding, the cells were transfected with 1300 ng gRNA plasmid using a Lipofectamine 3000 ratio of 1:0.8. Transduction of A549 and HeLa cells was performed 24h post seeding using 100 MOI AdV. The plasmid-transfected cells were harvested 48h post transfection. The AdV-transduced cells were harvested 72h post infection. This was followed by RNA isolation, Turbo-DNase digestion, reverse transcription, and qPCR. The qPCR experiments in panel a) and b) are shown as the mean of N = 2 technical replicates.*



Supplementary Figure S10: Extended mouse data. *The in vivo editing results plotted in main text Fig.5a show the arithmetic mean of five samples per animal. Specifically, the liver of each mouse was separated into its five lobes and the editing yields were determined for each liver lobe individually via Sanger sequencing. Here, the five RNA sequencing results (mean ± s.d.) for each animal are given in full detail, together with the arithmetic mean plotted in Fig. 5a.*



Supplementary Figure S11: qPCR quantification of hepatic ADAR expression levels in untreated C57BL/6 mice compared to mice treated with reporter and CLUSTER guide RNA via hydrodynamic tail-vein injection. *The hepatic total ADAR1, specific ADAR1p150 and ADAR2 expression was measured in technical triplicate in 8 untreated and 8 treated mice. Data are shown as the mean ± s.d. of N = 8 animals. The RNA used for the "No HDTV" measurement derived from 3 male and 5 female untreated mice. The RNA used for the "HDTV [Reporter+guide RNA]" measurement derived from the editing group 15-13-11-20p8 (3 mice, all female) and the editing group 20-15-15-20p8 (5 mice, 3 male, 2 female) that gave the results shown in Fig. 5a. The primer pairs applied in qPCR were characterized in Fig. S15. The measured ADAR Ct values were normalized to the geometric mean of the murine Rps29 and the murine ACTB housekeeping genes. The fold change was determined by means of the ΔΔC(t) method. For statistical analysis, a Mann-Whitney U-Test (Two-tailed, non-parametric) was applied. There is no significant difference in total ADAR expression between the HDTV treated and the untreated mice. In particular, no induction of ADAR1p150 was detectable.*



Supplementary Figure S12: The expression of endogenous ADAR1 in HeLa cells was unaffected by transfection of CLUSTER or LEAPER guide RNAs. *Western blot analysis of ADAR1 expression for editing the BMPR2 mutation W298X amber (delivered as a cDNA on plasmid), performed under the same conditions as the editing experiment shown in Extended Data Figure 5, but scaled up to 6-well scale. 10 µg protein-lysate were loaded per lane. The anti-ADAR1 antibody was used in a 1:250 dilution. The anti-β-actin antibody was used in a 1:5000 dilution. Both images were taken with 2 min. exposure time. No further image processing with respect to contrast or brightness was done. Clearly, there is neither ADAR1p150 induction nor increased expression of ADAR1p110 protein in response to guide RNA expression. The molecular weight markers apply to both western blots. N = 1 experiment was performed.*

Stage	Temperature [°C]	Duration [s]	Cycle #
Initial denaturation	95	20	1
Denaturation	95	3	40
Annealing/Elongation	60	30	
Melt curve	95	15	1
	60	60	
	95	15	
	60	15	

Supplementary Figure S13: Thermocycler program applied during qPCR.

Calculation of ΔΔC(t) for the target gene between two different samples:

$\Delta C(t)_{\text{treatment A (e.g. plasmid)}}$ = Treatment A mean $C(t)_{\text{target gene (e.g. R/G gRNA)}}$
– Treatment A mean $C(t)_{\text{reference gene (e.g. U6 snRNA)}}$

$\Delta C(t)_{\text{treatment B (e.g. AdV)}}$ = Treatment B mean $C(t)_{\text{target gene (e.g. R/G gRNA)}}$
– Treatment B mean $C(t)_{\text{reference gene (e.g. U6 snRNA)}}$

$\Delta\Delta C(t)$ = $\Delta C(t)_{\text{treatment A (e.g. plasmid)}} - \Delta C(t)_{\text{treatment B (e.g. AdV)}}$

Calculation of ΔΔC(t) between nucleus and cytoplasm for one reference gene:

$\Delta C(t)_{\text{nuclear}}$ = Nuclear fraction mean $C(t)_{\text{reference gene}}$

$\Delta C(t)_{\text{cytoplasmic}}$ = Cytoplasmic fraction mean $C(t)_{\text{reference gene}}$

$\Delta\Delta C(t)$ = $\Delta C(t)_{\text{nuclear}} - \Delta C(t)_{\text{cytoplasmic}}$

Calculation of ΔΔC(t) between nucleus and cytoplasm for one target gene:

$\Delta C(t)_{\text{nuclear}}$ = Nuclear fraction mean $C(t)_{\text{target gene (e.g. R/G gRNA)}}$
– Nuclear fraction geometric mean $C(t)_{\text{reference genes (e.g. HPRT1, GAPDH, U6 snRNA, Malat1)}}$

$\Delta C(t)_{\text{cytoplasmic}}$ = Cytoplasmic fraction mean $C(t)_{\text{target gene (e.g. R/G gRNA)}}$
– Cytoplasmic fraction geometric mean $C(t)_{\text{reference genes (e.g. HPRT1, GAPDH, U6 snRNA, Malat1)}}$

$\Delta\Delta C(t)$ = $\Delta C(t)_{\text{nuclear}} - \Delta C(t)_{\text{cytoplasmic}}$

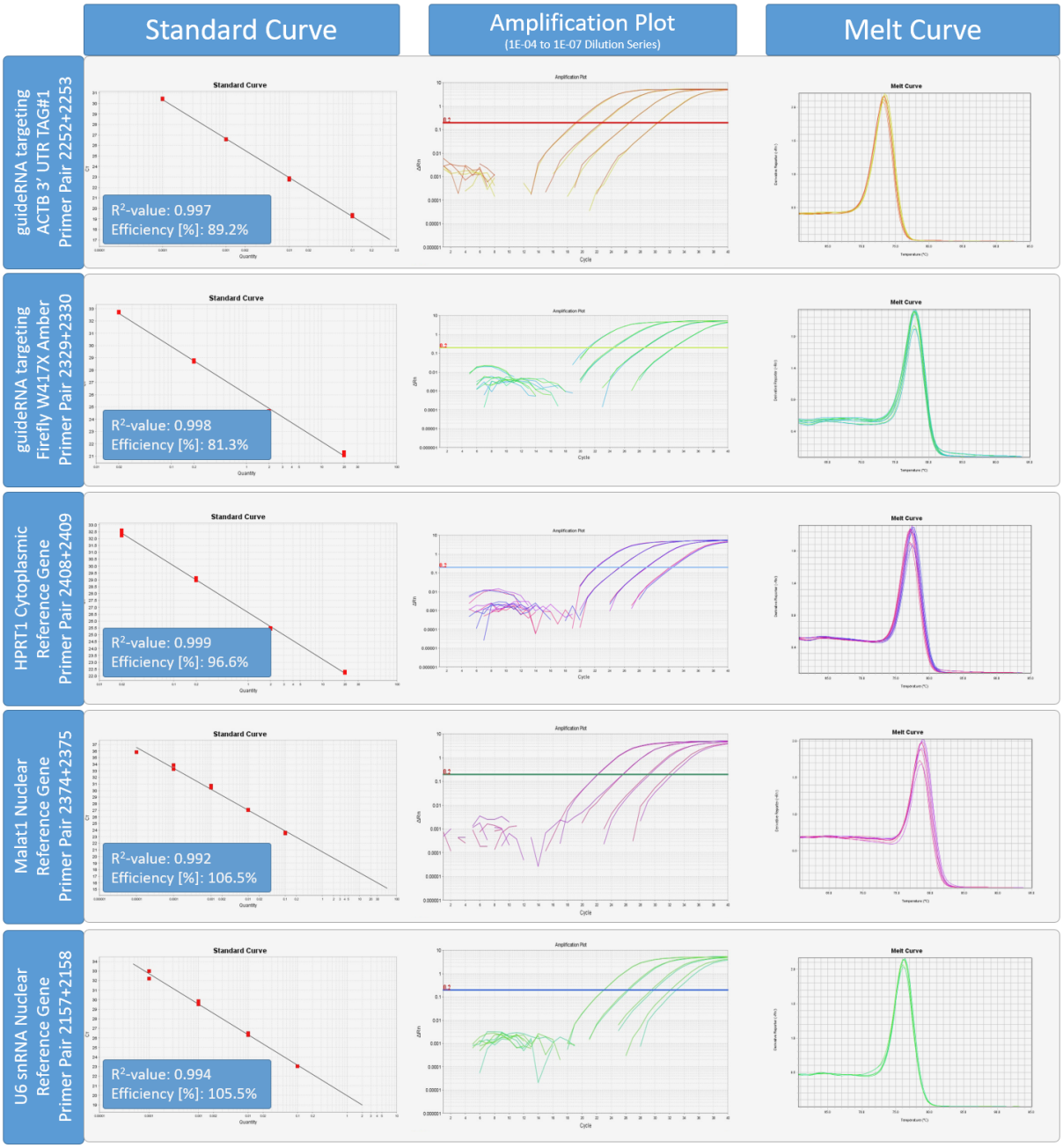
Calculation of fold change based on ΔΔC(t):

Fold change = $2^{-\Delta\Delta C(t)}$
Maximum fold change = $2^{-\Delta\Delta C(t) - \text{SD of target gene C(t)}}$
Minimum fold change = $2^{-\Delta\Delta C(t) + \text{SD of target gene C(t)}}$

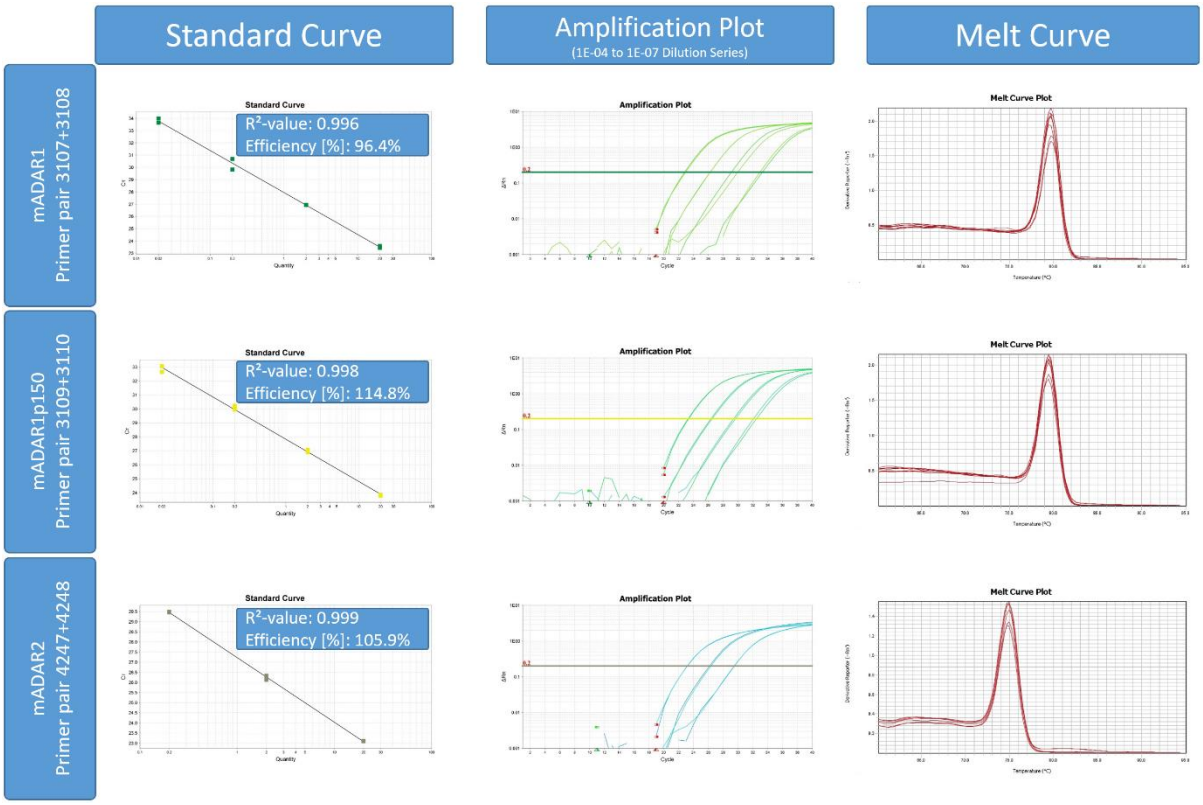
Calculation of the nuclear percentage for one target gene:

$$\%\ Nuclear\ localization = \frac{nuclear\ fold\ change}{(nuclear\ fold\ change\ + cytoplasmic\ fold\ change)} * 100$$

Supplementary Figure S14: ΔΔC(t) equations applied for analysis of qPCR data



Supplementary Figure S15: Optimization of qPCR conditions for guide RNA quantification. *All qPCR primer pairs used (see Supplementary Table S2 & S3) were characterized for their efficiency and were in the acceptable range between 80% and 110%, with good standard curve fits represented by R^2-values >0.99. The dilution series (10^{-4}, 10^{-5}, 10^{-6}, 10^{-7}) resulted in a linear dynamic range up to dilutions of 10^{-7} in the amplification plots, and clean single-transition melting curves. The gRNA and GAPDH primer pairs were established in our lab. The primer pair for HPRT1 came from L. Luan, L., et al. Sci Rep 2017, 7, 40050. The primer pair for Malat1 came from Q. Ji, et al. Br J Cancer 2014, 111, 736-48. The primer pair for U6 snRNA came from L.H. Yang, et al. PLoS One, 2014, 9, e115293. $2x10^5$ HEK-293T cells were transfected with 1300 ng plasmid, encoding either a 16p8 gRNA (16 nt long specificity domain, editing position 8, R/G motif version 1) targeting the β-actin 3' UTR, or targeting firefly-luciferase W417X amber. Lipofectamine 2000 was used in a 1:3 ratio. cDNA based on total RNA isolated from the β-actin gRNA sample was used for the β-actin guide RNA, HPRT1, Malat1 and U6 snRNA dilution series, while cDNA based on total RNA isolated from the firefly gRNA sample was used for the firefly-luciferase guide RNA dilution series.*



Supplementary Figure S16: Characterization of murine ADAR qPCR primer pairs. *All qPCR primer pairs used (see Supplementary Table S2 & S3) were characterized for their efficiency and were in the acceptable range between 80% and 120%, with good standard curve fits represented by R^2-values >0.99. The dilution series (10^{-4}, 10^{-5}, 10^{-6}, 10^{-7}) resulted in a linear dynamic range up to dilutions of 10^{-7} in the amplification plots, and clean single-transition melting curves. The ADAR1 and ADAR1p110 primer pairs were established in our lab. The ADAR2 primer pair came from Terajima, H., et al. (2017). "ADARB1 catalyzes circadian A-to-I editing and regulates RNA rhythm." Nat Genet 49(1): 146-151. For the dilution series a cDNA sample based on total RNA isolated from the liver of mouse #1 of the Fig. 5a editing group (15-13-11-20p8) was used.*

Supplementary Table S1: List of applied encoded guide RNAs and ASO (for the ease of use, the encoded guide RNAs are given as coding DNA sequences)

Sequence annotation:

- Targeting sequence
- Cluster of recruitment sequences
- R/G-motif V1
- R/G-motif V20
- R/G-motif V21
- BoxB-motif (with Xba-I attachment site)
- LEAPER targeting sequence
- mN = 2'O-methyl
- * = Phosphorothioate linkage

pTS #	Guide RNA Name	Sequence 5'->3'	Used in Figure
pTS946	Luciferase_W417X_RG-V21_0xRS_40p30	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCCCCAGTAGGCGATGTCGCCGCTGTGCAGCCAGCCGTCCTT	1b,c,e, S4, S5
pTS967	Luciferase_W417X_RG-V21_0xRS_20p8	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGT	1b,c,e,2a ,S4, S5
pTS970	Luciferase_W417X_RG-V21_8xRS(11-15-15-15-12-15-13-11-20p8)_3xA	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTAAACGCACAGCTCGAAAGTCCAAGTCCACCAAAAAGAAGGGCACCACCAAACTCCTCCTCGAAAAAAAAGCCGCAGATCAAAAACGAAGCTCTCGAAACCACAGCCACACCGAAAACACACCACGATCCGA	1b,c,e,f, 2a, S4, S5
pTS961	Luciferase_W417X_RG-V21_3xRS(15-13-11-20p8)_3xA	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTAAACGCACAGCTCGAAAGTCCAAGTCCACCAAAAAGAAGGGCACCACC	1b-e,g,h,2a, d,e,5a&b, EDF1-3, S4, S5, S7, S10, S11
pTS1017	Luciferase_W417X_NoRG_3xRS(15-13-11-20p8)_3xA	GTGCAGCCAGCCGTCCTTGTAAACGCACAGCTCGAAAGTCCAAGTCCACCAAAAAGAAGGGCACCACC	1d
pTS1181	Luciferase_W417X_RG-V21_1xRS_20p8	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTAAACGCACAGCTCG	2a
pTS1182	Luciferase_W417X_RG-V21_2xRS_20p8	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTAAACGCACAGCTCGAAAGTCCAAGTCCACC	2a
pTS1183	Luciferase_W417X_RG-V21_6xRS_20p8	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTAAACGCACAGCTCGAAAGTCCAAGTCCACCAAAAAGAAGGGCACCACCAAACTCCTCCTCGAAAAAAAAGCCGCAGATCAAAAACGAAGCTCTCG	2a
pTS1211	Luciferase_W417X_RG-V21_20xRS_20p8	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTAAACGCACAGCTCGAAAGTCCAAGTCCACCAAAAAGAAGGGCACCACCAAACTCCTCCTCGAAAAAAAAGCCGCAGATCAAAAACGAAGCTCTCGAAACCACAGCCACACCGAAAACACACCACGATCCGAAAGAACGCTCATCTCAAACTCGCCGGCGGTCCCAAACTCCTCCACGTCTCCAAAATCAACGAACGAAAAGAAAAGAATCCAAAACAACGTCAGGAAATCTCCAAAACCAAAATCAATCACATCAAACACCCCAATCAAAAACACCGCGCAAAGAATCAAGAACAAAAGCCCACCACT	2a
pTS1261	Luciferase_W417X_RG-V21_3xRS(15-L-15-S-15-S-20p8)	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTAAAACGCACAGCTCGCCGAAAGTCCAAGTCCACCACAAACGAAAGCCGCAGATC	2b

pTS1262	Luciferase_W417X_RG-V21_3xRS(15-L-15-L-15-S-20p8)	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTAAAACGCACAGCTCGCCGAAACGAAAGCCGCAGATCAAACCACAGCCACACCGA	2b
pTS1263	Luciferase_W417X_RG-V21_3xRS(15-S-15-L-15-S-20p8)	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTAAAACGCACAGCTCGCCGAAACGAAAGCCGCAGATCAAAGAACATGCCGAAGCC	2b
pTS1264	Luciferase_W417X_RG-V21_3xRS(15-S-15-S-15-L-20p8)	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTAAACGAAAGCCGCAGATCAAAGAACATGCCGAAGCCAAAATGGGGTCGCGGGCA	2b
pTS1265	Luciferase_W417X_RG-V21_3xRS(15-S-15-L-15-L-20p8)	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTAAACGAAAGCCGCAGATCAAACCACAGCCACACCGAAAACACCCAACACGGGCA	2b
pTS1266	Luciferase_W417X_RG-V21_3xRS(15-L-15-S-15-L-20p8)	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTAAACGAAAGCCGCAGATCAAAGAACATGCCGAAGCCAAACACCCAACACGGGCA	2b
pTS1267	Luciferase_W417X_RG-V21_3xRS(15-L-15-L-15-L-20p8)	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTAAACGAAAGCCGCAGATCAAACCACAGCCACACCGAAAACTCGCTCAACGAACG	2b,c
pTS1259	Luciferase_W417X_RG-V21_3xRS(15-S-20-S-15-S-20p8)	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTAAAACGCACAGCTCGCCGAAACCGGTGTCCAAGTCCACCACAAAAAGAAGGGCACCACC	2c
pTS1260	Luciferase_W417X_RG-V21_3xRS(15-S-15-S-20-S-20p8)	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTAAACACGGACGCACAGCTCGCCGAAAGTCCAAGTCCACCACAAAAAGAAGGGCACCACC	2c
pTS1277	Luciferase_W417X_RG-V21_3xRS(20-L-15-L-15-L-20p8)	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTAAACGAAAGCCGCAGATCAAACCACAGCCACACCGAAAAGAGAACTCGCTCAACGAACG	2c
pTS1278	Luciferase_W417X_RG-V21_3xRS(15-L-20-L-15-L-20p8)	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTAAACGAAAGCCGCAGATCAAAGGGGCCACAGCCACACCGATAAACTCGCTCAACGAACG	2c
pTS1279	Luciferase_W417X_RG-V21_3xRS(15-L-15-L-20-L-20p8)	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTAAACGACCCGAAAGCCGCAGATCAAACCACAGCCACACCGAAAACTCGCTCAACGAACG	2c
pTS1188	Luciferase_W417X_RG-V21_3xRS(15-S-15-S-15-S-20p8)	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTAAAACGCACAGCTCGCCGAAAGTCCAAGTCCACCACAAAAAGAAGGGCACCACC	2c,b,e,f, EDF4
pTS1258	Luciferase_W417X_RG-V21_3xRS(20-S-15-S-15-S-20p8)	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTAAAACGCACAGCTCGCCGAAAGTCCAAGTCCACCACAAACCTCGAAGAAGGGCACCACC	2c,f, 5a,b S10, S11
pTS1090	Luciferase_W417X_RG-V21_3xRS-25p8_3xA	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTCGATGAAAACGCACAGCTCGAAAGTCCAAGTCCACCACCAAAAGAAGGGCACC	2d
pTS1091	Luciferase_W417X_RG-V21_3xRS-30p8_3xA	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTCGATGAGAGCAAAACGCACAGCTCGCCGAAAGTCCAAGTCCACAAAAGAAGGGCACCA	2d
pTS1092	Luciferase_W417X_RG-V21_3xRS-35p8_3xA	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTCGATGAGAGCGTTTGAAACACAGCTCGCCGAAATCCAAGTCCACCAAAAAGAAGGGCACCACC	2d
pTS1093	Luciferase_W417X_RG-V21_3xRS-40p8_3xA	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTCGATGAGAGCGTTTGTAGCCAAAGGACGCACAGCAAATCCAAGTCCACCACAAAAAGAAGGGCACCAC	2d
pTS1100	Luciferase_W417X_RG-V21_3xRS-25p13_3xA	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCCCGCTGTGCAGCCAGCCGTCCTTGTAAAACGCACAGCTCGCCGAAAGTCCAAGTCCACCAAAAAGAAGGGCACCA	2d

pTS1101	Luciferase_W417X_RG-V21_3xRS-30p18_3xA	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCTGTCGCCGCTGTGCAGCCAGCCGTCCTTGTAAAACGCACAGCTCAAATCCAAGTCCACCACCAAAAGAAGGGCACCAC	2d
pTS1102	Luciferase_W417X_RG-V21_3xRS-35p23_3xA	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGGCGATGTCGCCGCTGTGCAGCCAGCCGTCCTTGTAAAACGCACAGCTCGCCGCAAATCCAAGTCCACCAAAAAGAAGGGCACCA	2d
pTS1103	Luciferase_W417X_RG-V21_3xRS-40p30_3xA	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCCCCAGTAGGCGATGTCGCCGCTGTGCAGCCAGCCGTCCTTAAAACGCACAGCTCGCCGAAATCCAAGTCCACCAAAAGAAGAAGGGCAC	2d
pTS1184	Luciferase_W417X_RG-V21_3xRS(7-7-7-20p8)	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTAAACCTCGGGAAAAAGTCCAAAAGAAGAAG	2e
pTS1185	Luciferase_W417X_RG-V21_3xRS(9-9-9-20p8)	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTAAAACGCACAGCAAACACACCCAGAAATCCACCACC	2e
pTS1186	Luciferase_W417X_RG-V21_3xRS(11-11-11-20p8)	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTAAACGCACAGCTCGAAACAAGTCCACCAAAAAGGGCACCACC	2e
pTS1187	Luciferase_W417X_RG-V21_3xRS(13-13-13-20p8)	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTAAAACGGACGCACAGCAAACCAAGTCCACCACAAAAGAAGGGCACCAC	2e
pTS1190	Luciferase_W417X_111p56_LEAPER	CACGATGAAGAAGTGCTCGTCCTCGTCCCAGTAGGCGATGTCGCCGCTGTGCAGCCAGCCGTCCTTGTCGATGAGAGCGTTTGTAGCCTCGGGGTTGTTAACGTAGCCGCT	2f
pTS1129	AHI_W725X_RG-V21_9xRS_20p8	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCAACTTTCCATATCCGTATCAAAAAAACAAAAGAAGGAAAAAAACGTCCAGAAGAAAGAGAAATCAAGAAAAGAAGAGCCATCAAAACCTCAACAACACAAAAAGAAATCAAGAATAAACCTCGAAGCAAAAAAGAAACAGGCCGTCCACAAAGGTCATCATCAAGCAA	3a
pTS1131	COL3A1_W1278X_RG-V21_9xRS_20p8	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTCAACCCAGTATTCTCCACAAAAGAACCATCAGAAAGGGCAAAACCGCAAAACCTCCAATCCCAGCAAAACAACACCACCACAGCAAAACCCTCAGATCCTCAAAGGTCCAACAGGTCCTCAAACAGCAGCTCCACGAAAGGCCAGCAGGGCCACAAACACCACGATCA	3a
pTS1132	FANCC_W506X_RG-V21_9xRS_20p8	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGACATCCCAGGCGATCGTGTAAAGCCTCCCATCACAAATCTCAGCCCATCCTCCAAAGAACCAGCTCTCAAAGAAACAAACCGCAGCAAATCCAGGGCCCCATCGAAACTCATCAACAACCCGGAAAGCCGAAGCCAGAAAACCTCAAAGAACTCAAAGGACACAAACTC	3a
pTS1135	IL2RG_W237X_RG-V21_6xRS_20p8	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCTTCACTCCAATGCTGAGCACAAAGCTCCGAACACGAAACAAAGCCCATCCACACAAAGTCCCAGTCAGTCCGAAAACAGCCAGAAGAAAGAACACAAAACACAAACAGCCCCACTCCC	3a
pTS1136	MYBPC3_W1098X_RG-V21_9xRS_20p8	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTACCCCCAGAGCTCCGTGTAAATCCACTCCAGAAAACGTCAGTCACCCGAAAGCACCGTCACCAAATCACCTCCTCGAAATCACAGGCTCCCCGACAAAGAAAAGCAGCCGAAACAGCCACCCACTAAATCCCACGCGCTAAAAGACGCGCATCT	3a
pTS1180	PINK1_W437X_RG-V21_3xRS_20p8	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCCACTGCCCAGGCATCAGCCTAAACCGCCCCGATCCACGAAATCTCATCAGCCAGAAAATCACCAGCCA	3a

pTS1130	BMPR2_W298X_RG-V21_9xRS_20p8	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGCTTACCCAGTCACTTGTGTAAAACTCCATCACAAAAACTCTCTCATCTCCAACAAATCCATCAAAGGCACTCAAAGCAAAGGAAAACACAAAGACGCTCATCCAAGAAAACCTCGGCCAATCAGAAAATCAAAACAGCAAAAACAGCAGAAACGAAAAGACCAACATCC	3a, EDF5, S12
pTS1343	GAPDH_L157L_RG-V21_3xRS(18-15-15-20p8)_3xA	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGGGGTGCCAAGCAGTTGGTGAAAGGGGCATCAGCAGAGAAAGCGCCAGCATCGCCCAAATGGAGGGATCTCGCTCCT	3c
pTS1345	GAPDH_L157L_RG-V21_3xRS(20-20-20-20p8)_3xA	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGGGGTGCCAAGCAGTTGGTGAAAGGGGCATCAGCAGAGGGGAAAACTCAGCGCCAGCATCGCCCAAACGAACAGGAGGAGCAGAGAG	3c
pTS1347	GPI_V627V_RG-V21_3xRS(20-15-15-20p8)_3xA	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCTGCCGTCCACCAGGATGGGTAAACACGGCTCGACCCTCAAAGAACATCCGCTCCCGAAACGCATCACGTCCTCCGTCAC	3c
pTS1348	GPI_V627V_RG-V21_3xRS(20-20-20-20p8)_3xA	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCTGCCGTCCACCAGGATGGGTAAAGCAGCACGGCTCGACCCTCGAAACCGGGCGGCCTCCACGCCCCAAAATCACGTCCTCCGTCACCAG	3c
pTS1351	PAICS_V269_RG-V21_3xRS(19-15-15-20p8)_3xA	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCTCAACACCACAACCCTGCACAAAGCAACCCACTCAAAGAAAAGTCTCCAGGAATCAAAATAACATCAGCAAGAACAAT	3c
pTS1352	PAICS_V269V_RG-V21_3xRS(20-20-20-20p8)_3xA	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCTCAACACCACAACCCTGCACAAAGGCCAGAGTCTCCAGGAATCAAAAGAAGTCCAGCAAAGCAAAAAAACCGCGCCGGGAAGCGAGGCA	3c
pTS1353	PDE4D_L291L_RG-V21_3xRS(19-15-15-20p8)_3xA	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGGGTCTCCAGCTGGTCCAGAAAAGCTCCTCCAGGGTCTAAAAATCAAGTCATCTCCAAATGTCCACATCAAAACGCCT	3c
pTS1354	PDE4D_L291L_RG-V21_3xRS(20-20-20-20p8)_3xA	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGGGTCTCCAGCTGGTCCAGAAAACAGCTCCTCCAGGGTCTCGCAAATCACAATCAAGTCATCTCCGAAAAAGGGACTCCGTCCCGCAGA	3c
pTS1511	GPI_V627V_20p8_0xRC_RG-V21	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCTGCCGTCCACCAGGATGGGT	3c
pTS1512	NUP43_V233V_20p8_0xRC_RG-V21	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCCAGTAGCCACAACATGCTGT	3c
pTS1561	GUSB_L456L_0xRC	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCCAGATTCCAGGTGGGACGCA	3c
pTS1562	GAPDH_L157L_0xRC	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGGGGTGCCAAGCAGTTGGTG	3c
pTS1563	PAICS_V269V_0xRC	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCTCAACACCACAACCCTGCAC	3c
pTS1564	PDE4D_L291L_0xRC	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGGGTCTCCAGCTGGTCCAGA	3c
pTS1357	NUP43_V223V_RG-V21_3xRS(20-20-20-20p8)_3xA	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCCAGTGGCCACAACATGCTGTAAAGGCACTCGGTCACCAGTCAGAAAAAACCCTCCATCAGAGTCCAAAAAGGCAGCGGTCGCCAGCGGG	3c, 4a
pTS1358	NUP43_V223V_RG-V21_3xRS(20-15-15-20p8)_3xA	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCCAGTGGCCACAACATGCTGTAAACGGTCACCAGTCAGTAAAAACCCTCCATCAGAGAAAAGGCAGCGGTCGCCAGCGGG	3c, 4a
pTS1360	GusB_L456L_RG-V21_3xRS(20-20-20-20p8)_3xA	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCCAGATTCCAGGTGGGACGCAAAAGCCACAGACCACATCACGACAAACA	3c, 4b

		CGCCGGGACACTCATCGAAAAGCACCAAGCCAGCGAAGCAG	
pTS1361	GusB_L456L_RG-V21_3xRS(20-15-15-20p8)_3xA	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCCAGATTCCAGGTGGGACGCAAAAAGACCACATCACGACAAACACGCCGGGACACTCAAAGCACCAAGCCAGCGAAGCAG	3c, 4b
pTS1341	β-Actin_3'UTR_TAG#1_RG-V21_3xRS(20-15-15-20p8)_3xA	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCACGCAACCAAGTCATAGTCCAAAGGAGGGGCCGGACTCAAACCGCCGATCCACACGAAATGCGCTCAGGAGGAGCAATG	3d
pTS1342	β-Actin_3'UTR_TAG#1_RG-V21_3xRS(20-20-20-20p8)_3xA	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCACGCAACCAAGTCATAGTCCAAAAGCCGCCGATCCACACGGAGAAACCTCAGGGCAGCGGAACCGCAAAGCTCGAAGTCCAGGGCGACG	3d
pTS1344	GAPDH_3'UTR_TAG#1_RG-V21_3xRS(20-15-15-20p8)_3xA	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCAGGGGTCCACATGGCAACTGAAACAGGGACTCCCCAGCAAAGAAGTCAGAGGAGACAAATAGACGGCAGGTCAGGTCCA	3d
pTS1346	GAPDH_3'UTR_TAG#1_RG-V21_3xRS(20-20-20-20p8)_3xA	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCAGGGGTCCACATGGCAACTGAAATGTGGCAGGGACTCCCCAGCAAAGAAGTCAGAGGAGACCACCTAAATAGACGGCAGGTCAGGTCCA	3d
pTS1565	RAB7A_3'UTR_TAG#1_0xRC	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCCTGCCGCCAGCTGGATTTCC	3d
pTS1566	β-Actin_3'UTR_TAG#1_0xRC	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCACGCAACCAAGTCATAGTCC	3d
pTS1567	GAPDH_3'UTR_0xRC	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCAGGGGTCCACATGGCAACTG	3d
pTS1362	RAB7A_3'UTR_TAG#1_RG-V21_3xRS(20-15-15-20p8)_3xA	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCCTGCCGCCAGCTGGATTTCCAAATTCAACCCTCCACCTAAATACAAAACTCAGCCTAAATAAAAAGACCACAAGACCAA	3d, 4c
pTS985	RAB7A_3'UTR_TAG#1_RG-V21_3xRS(19-13-11-20p8)_3xA	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCCTGCCGCCAGCTGGATTTCCAAATCAACCCTCCACCAAAACAAAACTCAGAAAAAAAAGACCACAAGACCAA	3d, 4c-f, EDF6-8
w/o #	hIDUA_W402X_RG-V21_3xRS(20-20p8-25-20)_3xA_ASO	mG*mG*mU*GUCGAGAAGAGGAGAACAAUAUGCUAAAUGUUGUUCUCGUCUCCUCGACACCGCCAGGACGCCCACCGUGUGAAACUGUCCAGGACGGUCCCGGCCUGCGAAAUUCGGCCCAGAGCUGCUC*C*U*A*A*A*G*U*G*G*G*G*U*C*G*U*U*G*C*C*C*A*G*mC*mC*mG	3g,h
pTS1194	NUP43_V233V_111p56_LEAPER	TAGTACCTTGTCTAACATCCCAAATACTCAACATTCCATCTTGGCCACCAGTAGCCACAACATGCTGTTGGTTGGGATGTCTATCAACACAGTGGAGTGGCACTCGGTCAC	4a
pTS1217	NUP43_V233V_111p56_LEAPER_AG6	TAGTACCTTGTCTAACATCCCAAAGACTCAACATTCCATCGTGGCCACCAGGAGCCACAACATGCTGGTGGGTGGGATGTCGATCAACACAGTGGAGTGGCACTCGGTCAC	4a
pTS1285	GusB_L456L_111p56_LEAPER	GGTCCAAGGATTTGGTGTGAGCGATCACCATCTTCAAGTAGTAGCCAGCAGATTCCAGGTGGGACGCAGGCTCGTTGGCCACAGACCACATCACGACCGCGGGGTGGTTCT	4b
pTS1216	RAB7A_3'UTR_TAG#1_111p56_LEAPER_AG20	GTCTTTGAGAAAAGGCGGACAGAAGTCTGGTGTCGACGGGACAGAAGACGGCCGCCAGCTGGAGGGCCCAAGGCGGAGGAACACTCGGCAATCCAAACAGGGGGCAACCCT	4c
pTS1193	RAB7A_3'UTR_TAG#1_111p56_LEAPER	GTCTTTGATAAAAGGCGTACATAATTCTTGTGTCTACTGTACAGAATACTGCCGCCAGCTGGATTTCCCAATTCTGAGTAACACTCTGCAATCCAAACAGGGTTCAACCCT	4c-f, S15, EDF7-8
pTS1418	Luciferase_W417X_RG-V21_3xRS(15-S-15-S-15-S-20p8)+mComp14-TripleHelix_wt	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTAAAACGCACAGCTCGCCGAAAGTCCAAGTCCACCACAAAAAGAAGGGCACCACCAAGCTTTTTCTTTTCCTGAGAAATTTCTCAGGTTTTGCTTTTTAAAAAAAAAGCAAAA	S8

pTS1419	Luciferase_W417X_RG-V21_3xRS(15-S-15-S-15-S-20p8)+mComp14-TripleHelix_variant_1	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTAAAACGCACAGCTCGCCGAAAGTCCAAGTCCACCACAAAAAGAAGGGCACCACCAAGCTTCTTTTCTTTTCCTGAGAAATTTCTCAGGTTTTGCTTTTCAAAAGAAAAGCAAAA	S8
pTS1420	Luciferase_W417X_RG-V21_3xRS(15-S-15-S-15-S-20p8)+mComp14-TripleHelix_variant_2	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTAAAACGCACAGCTCGCCGAAAGTCCAAGTCCACCACAAAAAGAAGGGCACCACCAAGCTTTCTTTCTTCTCCTGAGAAATTTCTCAGGTCTTGCTTTCTAAAAAGAAAGCAAGA	S8
pTS1421	Luciferase_W417X_RG-V21_3xRS(15-S-15-S-15-S-20p8)+mComp14-TripleHelix_wt+TRBR	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTAAAACGCACAGCTCGCCGAAAGTCCAAGTCCACCACAAAAAGAAGGGCACCACCAAGCTTTTTCTTTTAAAAGAGAAATTTCTCAAAATTTTGCTTTTTAAAAAAAAGCAAAA	S8
pTS1422	Luciferase_W417X_RG-V21_3xRS(15-S-15-S-15-S-20p8)+mComp14-TripleHelix_variant_1+TRBR	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTAAAACGCACAGCTCGCCGAAAGTCCAAGTCCACCACAAAAAGAAGGGCACCACCAAGCTTCTTTTCTTTTAAAAGAGAAATTTCTCAAAATTTTGCTTTTCAAAAGAAAAGCAAAA	S8
pTS1423	Luciferase_W417X_RG-V21_3xRS(15-S-15-S-15-S-20p8)+mComp14-TripleHelix_variant_2+TRBR	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTAAAACGCACAGCTCGCCGAAAGTCCAAGTCCACCACAAAAAGAAGGGCACCACCAAGCTTTCTTTCTTCTAAAAGAGAAATTTCTCAAAATCTTGCTTTCTAAAAAGAAAGCAAGA	S8
pTS1718	BMPR2_W298X_111p56_LEAPER	GTGAAGATAAGCCAGTCCTCTAGTAACAGAATGAGCAAGACGGCAAGAGCTTACCCAGTCACTTGTGTGGAGACTTAAATACTTGCATAAAGATCCATTGGGATAGTACTC	EDF5 & S12
pTS1691	Non-targeting_gRNA_RG-V21	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACC	EDF6, EDF8
pTS408	β-Actin_3'UTR_TAG#1_RG-V1_0xRS_16p8_BoxB	GTGGAATAGTATAACAATATGCTAAATGTTGTTATAGTATCCCACACGCAACCAAGTCATATCTAGAGGGCCCTGAAGAGGGCCC	S6b-e, S9
pTS558	Luciferase_W417X_RG-V20_0xRS_16p8_BoxB	GTGGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACCAGGTGCAGCCAGCCGTCCTCTAGAGGGCCCTGAAGAGGGCCC	S7
pTS1189	Luciferase_W417X_RG-V21_3xRS(RS#1-3mask_15-13-11-20p8)	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTAAACGCACAGCTCGAAAGTCCAAGTCCACCAAAAAGAAGGGCACCACCAAAGGTGGTGCCCTTCTTAAAGGTGGACTTGGACAAACGAGCTGTGCG	EDF2
pTS1295	Luciferase_W417X_RG-V21_3xRS(RS#3mask_15-13-11-20p8)	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTAAACGCACAGCTCGAAAGTCCAAGTCCACCAAAAAGAAGGGCACCACCAAAGGTGGTGCCCTTCTT	EDF2
pTS1296	Luciferase_W417X_RG-V21_3xRS(RS#2+3mask_15-13-11-20p8)	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTAAACGCACAGCTCGAAAGTCCAAGTCCACCAAAAAGAAGGGCACCACCAAAGGTGGTGCCCTTCTTAAAGGTGGACTTGGAC	EDF2
pTS1173	Luciferase_W417X_RG-V21_3xRS(15-13-11-20p8)_1xA	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTACGCACAGCTCGAGTCCAAGTCCACCAAAGAAGGGCACCACC	EDF3
pTS1174	Luciferase_W417X_RG-V21_3xRS(15-13-11-20p8)_2xA	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTAACGCACAGCTCGAAGTCCAAGTCCACCAAAAGAAGGGCACCACC	EDF3
pTS1175	Luciferase_W417X_RG-V21_3xRS(15-13-11-20p8)_4xA	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTAAAACGCACAGCTCGAAAAGTCCAAGTCCACCAAAAAAGAAGGGCACCACC	EDF3
pTS1176	Luciferase_W417X_RG-V21_3xRS(15-13-11-20p8)_5xA	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTAAAAACGCACAGCTCGAAAAAGTCCAAGTCCACCAAAAAAAGAAGGGCACCACC	EDF3

pTS1297	Luciferase_W417X_RG-V21_3xRS(15-13-11-20p8)_0xA	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTCGCACAGCTCGGTCCAAGTCCACCAAGAAGGGCACCACC	EDF3
pTS1298	Luciferase_W417X_RG-V21_3xRS(15-13-11-20p8)_10xA	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCGTGCAGCCAGCCGTCCTTGTAAAAAAAAAACGCACAGCTCGAAAAAAAAAAGTCCAAGTCCACCAAAAAAAAAAAAGAAGGGCACCAC	EDF3
pTS1402	Luciferase_W417X_RG-V21_3xRS(15-15-20p8-15)_3xA	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCTCGTCCTCGTCCCAGAAAGTGCAGCCAGCCGTCCTTGTAAAGTCCAAGTCCACCACAAAACGCACAGCTCGCCG	EDF4
pTS1403	Luciferase_W417X_RG-V21_3xRS(15-20p8-15-15)_3xA	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCTCAGCCGGTCCACGAAAATCGTCCTCGTCCCAGAAAGTGCAGCCAGCCGTCCTTGTAAAGTCCAAGTCCACCAC	EDF4
pTS1404	Luciferase_W417X_RG-V21_3xRS(20p8-15-15-15)_3xA	GGTGTCGAGAAGAGGAGAACAATATGCTAAATGTTGTTCTCGTCTCCTCGACACCCCCCGGCGTCGAAGAAAATCAGCCGGTCCACGAAAATCGTCCTCGTCCCAGAAAGTGCAGCCAGCCGTCCTTGT	EDF4

Supplementary Table S2: Primer List

Primer #	Primer Name	Sequence 5‘->3‘
w/o #	GAPDH_fw	caacagcctcaagatcatcag
w/o #	GAPDH_bw	ccttccacgataccaaagttg
w/o #	hIDUA_fw	atttacaacgacgaggcgga
w/o #	hIDUA_bw	tcatccgaccagaccagaac
w/o #	hIDUA_nested_fw	cgcgcttccaggtcaacaacac
w/o #	hIDUA_nested_bw	gtcgctcgcgtagatcagca
46	Luci_seq_mid_fw	ctattcttgcgcagcttgcaagac
144	BGH backward	ctagaaggcacagtcgaggc
196	PINK_Analyse_1fw	gagccaggagctggtcccagcgagccgag
278	PINK_R407Qfw	ggtacgtggatcagggcggaaacggc
335	Luci_Analyse_lvfw	ggtgtgaaccagcgcggcgagctg
1032	Beta-Aktin_fw	cagcagatgtggatcagcaagcaggag
1033	Beta-Aktin_bw	ggaagggggggcacgaaggctcatc
1159	GAPDH_ORF_seq_bw	gctgttgaagtcagaggagacc
1357	pShuttle-CMV-bw	gtggtatggctgattatgatcag
1758	GAPDH_5UTR_isoform1_fw	gccagccgagccacatcgc
1759	GAPDH_ORF2_seq_fw	gattccacccatggcaaattccatg
2157	qPCR_snRNA-U6_fw	gcttcggcagcacatatactaaaat
2158	qPCR_snRNA-U6_bw	cgcttcacgaatttgcgtgtcat
2252	β-Actin_qPCR-AdapterD1+guide+RG_bw	gaaatcgtcgctgactatcgtatgacttggttgcgtgtgggatac
2253	β-Actin_qPCR-AdapterD2+RG-V1_bw	cagctgtaccgttgaatcgagtggaatagtataacaatatgctaa
2329	FLuc_qPCR-AdapterD1+guide+RG-V20_bw	gactatcgggacggctggctgcacgtggtcgag
2330	qPCR-AdapterD2+RG-V20_bw	cagctgtaccgttgaatcgagtggtcgagaagaggagaacaatat
2374	Malat1_qPCR_fw	aggcgttgtgcgtagagga
2375	Malat1_qPCR_bw	ggatttttaccaaccactcgc
2408	HPRT1_V2_qPCR_fw	tggcgtcgtgattagtgatg
2409	HPRT1_V2_qPCR_bw	acccttttccaaatcctcagc
2716	ADAR1_fw	gcatttgaggatggactacg
2717	ADAR1_bw	tccttagtcttcccggattg
2898	Fluc_W417X_Universal_fw	acttggacaccggtaagacac
2899	Fluc_W417X_Universal_bw	acgatgaagaagtgctcgtcc
2901	AHI_W725X_Amber_fw	gccgtatcccaaacaaacacc
2904	BMPR2_W298X_Amber_seq_bw	acctcgcttatggctgcatt
2905	COL3A1_W1278X_Amber_fw	tgctgggattggaggtgaaaa
2907	FANCC_W506X_Amber_fw	attcacttcggaggatgggc
2913	IL2RG_W237X_Amber_fw	aaagtccagaagtgcagcca
2915	MYBPC3_W1098X_Amber_fw	gcgtgcattcaggcacttac
2917	GPI_V627V_fw	gtgtaccttctagtcccgcc
2918	GPI_V627V_bw	gccaatgttgatgacgtccg
2919	GusB_L456L_fw	caacaagcatgaggatgcgg
2920	GusB_L456L_bw	gtgcccgtagtcgtgatacc
2934	NUP43_V233V_bw	gtactgcttcttcctccttggtg
2935	PAICS_V269V_fw	gagaaatcctggttgcccca
2936	PAICS_V269V_bw	ttacttctgcctgccactgc

2939	RAB7A_3'UTR_TAG#1+2_fw	gccccattacaggctcacac
2940	RAB7A_3'UTR_TAG#1+2_bw	ttgaagtgtggagcaggggg
2943	GAPDH_3'UTR_TAG#2_fw	cctgacctgccgtctagaaa
2944	GAPDH_3'UTR_TAG#2_bw	tggtacatgacaaggtgcgg
3015	PDE4D_#3_L291L_fw	accataacagaggaggcctac
3016	PDE4D_#3_L291L_bw	gataacagtcaagggccggt
3041	RAB7A_Exon5_fw	ccagacgattgcacggaatg
3107	Murine_ADAR1_fw	tcccgccattaccctgtctt
3108	Murine_ADAR1_bw	catggtacggagtctttcccc
3109	Murine_ADAR1p150_fw	cttgccggcactatgtctca
3110	Murine_ADAR1p150_bw	ctgcgggtatctccacttgc
3206	GusB_L456L_Seq_NonRCsite_fw	atgcaggtgatggaagaag
3207	Nup43_V233V_Seq_NonRCsite_fw	cagacaacaaggaaatgagc
3236	Murine_ACTB_fw	gagcgcaagtactctgtgtg
3237	Murine_ACTB_bw	aaacgcagctcagtaacagtc
3445	Murine_IDUA_fw	cccacttggatgcattcttggacctt
3446	Murine_IDUA_bw	gcgtgggtgtcatcactagtgtag
3454	NUP43_5'UTR_fw	ctgctgcggccgctttcg
3590	RAB7A_RC_seq_fw	tgatatggagttggcattgg
3850	Fluc_W417X_Nested_fw	tgggtgtgaaccagcgcg
3851	Fluc_W417X_Nested_bw	gtcccagtaggcgatgtcg
4242	Murine_Rps29_fw	tgaaggcaagatgggtcac
4243	Murine_Rps29_bw	gcacatgttcagcccgtatt
4247	Murine_ADAR2_fw	ttgccctgaaggagttttg
4248	Murine_ADAR2_bw	gagggcttcttgactggc
4249	BMPR2_W298X_Amber_fw	tcaagaacggctatgtgcgt
4250	BMPR2_W298X_Amber_seq2_bw	tgctccatatcgacctcggcc

Supplementary Table S3: Primer pair & Sequencing primer list

Target	Primer fw & bw	Sequencing Primer
ACTB_3'UTR_TAG#1_gRNA_qPCR	2252+2253	None
ADAR1_qPCR	2716+2717	None
AHI_W725X (pTS1159)	2901 & 144	2901
Beta-Actin_3'UTR_TAG#3	1032+1033	1032
BMPR2_W298X (pTS1160)	4249 & 144	2904 & 4250
COL3A1_W1278X (pTS1161)	2905 & 144	2905
Dual-Luciferase_W417X_(In-Vitro)_(pTS554 & pTS555)	46+1357	335
Dual-Luciferase_W417X_(In-Vivo Nested)	3850+3851	3850
Dual-Luciferase_W417X_(In-Vivo)_(pTS656 & pTS657)	2898+2899	Sequencing after nested PCR
FANCC_W506X (pTS1162)	2907 & 144	2907
Firefly _W417X_Amber_gRNA_qPCR	2329+2330	None
GAPDH_3'UTR_TAG#2	2943+2944	2943
GAPDH_L157L	1159+1758	1759
GAPDH_qPCR	GAPDH_fw+ GAPDH_bw	None
GPI_V627V	2917+2918	2917
GusB_L456L	2919+2920	2919 (RS's), 3206 (Target site)
HPRT1_qPCR	2408+2409	None
Human_IDUA	hIDUA_fw+ hIDUA_bw	Sequencing after nested PCR
Human_IDUA_nested	hIDUA_nested_fw+ hIDUA_nested_bw	hIDUA_nested_bw
IL2RG_W237X (pTS1165)	2913 & 144	2913
Malat1_qPCR	2374+2375	None
Murine_ACTB_qPCR	3236+3237	None
Murine_ADAR1_qPCR	3107+3108	None
Murine_ADAR1p150_qPCR	3109+3110	None
Murine_ADAR2_qPCR	3247+3248	None
Murine_IDUA	3445+3446	3446
Murine_Rps29_qPCR	4342+4343	None
MYBPC3_W1098X (pTS1166)	2915 & 144	2915
NUP43_V233V	3454+2934	3454 (RS's), 3207 (Target site)
PAICS_V269V	2935+2936	2935
PDE4D_L291L	3015+3016	3015
PINK1_W437X (pTS65)	196 & 144	278
RAB7A_3'UTR_TAG#1	3041+2940	3590 (RS's), 2939 (Target site)
U6_snRNA_qPCR	2157+2158	None

Supplementary Table S4: Sense oligo list

Primer #	Sense Oligo Name	Sequence 5‘->3‘
3115	pTS1129 sense oligo	tgatttctttttgtgttgttgaggttttgatggctcttctttttcttgatttctctttcttctggacgttttttttccttctttttgttttttgatacg gatatggaaacac
3116	pTS1130 sense oligo	tttcttggatgagcgtctttgtgttttcctttgctttgagtgcctttgatggatttgttggagatgagagagttttttgtgatggagtt ttacacaagtgactgggtatag
3117	pTS1131 sense oligo	gacctttgaggatctgagggttttgctgtggtggtgttgttttgctgggattggaggttttgcggttttgccctttctgatggttc ttttgtggagaatactgggttccg
3118	pTS1132 sense oligo	agtttcgatggggccctggatttgctgcggtttgtttctttgagagctggttctttggaggatgggctgagatttgtgatggg aggctttacacgatcgcctgggatcgg
3121	pTS1135 sense oligo	ggctgtttgtgttttgtgttctttcttctggctgttttcggactgactgggactttgtgtggatgggctttgtttcgtgttcggagctt tgtgctcagcattggagtcct
3122	pTS1136 sense oligo	gctgcttttctttgtcggggagcctgtgatttcgaggaggtgatttggtgacggtgctttcgggtgactgacgttttctggag tggatttacacggagctctgggggtc
3125	pTS1180 sense oligo	tggctggtgattttctggctgatgagatttcgtggatcggggcggtttaggctgatgcctgggcaccc
3129	pTS1194 sense oligo	gaccgagtgccactccactgtgttgatagacatcccaaccaacagcatgttgtggctactggtggccaagatggaat gttgagtatttgggatgttagacaaggtacgcc
3133	pTS985 sense oligo	ttggtcttgtggtctttttttctgagtttgttttggtggagggttgatttggaaatccagctggcgggcc
3134	pTS1193 sense oligo	ggttgaaccctgtttggattgcagagtgttactcagaattgggaaatccagctggcggcagtattctgtacagtagaca caagaattatgtacgccttttatcaaagtga
3135	pTS961 sense oligo	ggtggtgcccttctttttggtggacttggactttcgagctgtgcgtttacaaggacggctggctggtg
3293	pTS1285 sense oligo	agcttcccaaaaaaagaaccaccccgcggtcgtgatgtggtctgtggccaacgagcctgcgtcccacctggaatct gctggctactactgaagatggtgatcgctcacaccaaatccttggaccg
3568	pTS1341 sense oligo	cattgctcctcctgagcgcatttcgtgtggatcggcggtttgagtccggccccctcctttggactatgacttcgttggca
3569	pTS1342 sense oligo	cgtcgccctggacttcgagctttgcggttccgctgccctgaggtttctccgtgtggatcggcggcttttggactatgacttc gttggca
3570	pTS1343 sense oligo	aggagcgagatccctccatttgggcgatgctggcgctttctctgctgatgcccctttcaccaactgcttcgcacggt
3571	pTS1344 sense oligo	ctctctgctcctcctgttcgtttgggcgatgctggcgctgagttttcccccctctgctgatgcccccttttcaccaactgcttcg cacggt
3572	pTS1345 sense oligo	tggacctgacctgccgtctatttgtctcctctgacttctttgctggggagtccctgtttcagttgccatgtcgaccgga
3573	pTS1346 sense oligo	tggacctgacctgccgtctatttaggtggtctcctctgacttctttgctggggagtccctgccacatttcagttgccatgtcg accgga
3574	pTS1347 sense oligo	gtgacggaggacgtgatgcgtttcgggagcggatgttctttgagggtcgagccgtgtttacccatcctggtcgacgcg g
3575	pTS1348 sense oligo	ctggtgacggaggacgtgattttggggcgtggaggccgcccggtttcgagggtcgagccgtgctgctttacccatcct ggtcgacgcgg
3578	pTS1351 sense oligo	attgttcttgctgatgttattttgattcctggagacttttctttgagtgggttgctttgtgcagggttgtcgtgtacg
3579	pTS1352 sense oligo	tgcctcgcttcccggcgcggtttttttgctttgctggacttctttgattcctggagactctggcctttgtgcagggttgtcgtgt acg
3580	pTS1353 sense oligo	aggcgttttgatgtggacatttggagatgacttgatttttagaccctggaggagcttttctggaccagctcgagaggt
3581	pTS1354 sense oligo	tctgcgggacggagtccctttttcggagatgacttgattgtgatttgcgagaccctggaggagctgttttctggaccagc tcgagaggt
3583	pTS1357 sense oligo	cccgctggcgaccgctgccttttggactctgatggagggtttttctgactggtgaccgagtgccttacagcatgttgtc gccagcc
3584	pTS1358 sense oligo	cccgctggcgaccgctgccttttctctgatggagggtttttactgactggtgaccgtttacagcatgttgtcgccagcc
3586	pTS1360 sense oligo	ctgcttcgctggcttggtgcttttcgatgagtgtcccggcgtgtttgtcgtgatgtggtctgtggcttttgcgtcccacctcga atgcc
3587	pTS1361 sense oligo	ctgcttcgctggcttggtgctttgagtgtcccggcgtgtttgtcgtgatgtggtctttttgcgtcccacctcgaatgcc
3588	pTS1362 sense oligo	ttggtcttgtggtcttttttatttaggctgagttttgtatttaggtggagggttgaatttggaaatccagctcgcgggcc
3589	pTS1258 sense oligo	ggtggtgcccttcttcgaggtttgtggtggacttggactttcggcgagctgtgcgttttacaaggacggctcgctggtg

Supplementary Table S5: List of the antibodies used to generate the western blot illustrated in Fig. S3 & S12.

Antibody	Target Protein	Produced in	Immunoglobulin Class	Dilution used	Supplier	Order #	Against	Validation
ADAR1 (15.8.6)	α-ADAR1	Mouse	monoclonal IgG	1:250 or 1:1000	Santa Cruz	sc-73408	amino acids 440-826 corresponding to the middle region of ADAR1 of human origin	Validated in our lab via siRNA KO and Western Blot PMID: # 28669490 PMID: # 28278381 PMID: # 27573237
Clone AC-15	α-Beta-Actin	Mouse	monoclonal IgG	1:40.000	Sigma Aldrich	A5441	Actin N-terminal peptide, Ac-Asp-Asp-Asp-Ile-Ala-Ala-Leu-Val-Ile-Asp-Asn-Gly-Ser-Gly-Lys	PMID: # 15809369 PMID: # 15048076 PMID: # 21217779
Anti-mouse–HRP	α-Mouse IgG	Goat	Polyclonal IgG	1:5.000	Jackson Immuno Research Laboratories	115-035-003	whole molecule mouse IgG	Validated in our lab via Western Blot (in several cell lines) two examples of the 1181 citations registered on September 9th 2021: PMID: # 32319599 PMID: # 32782498

Supplementary Notes:

Table of supplementary notes

1. Cloning strategy for CLUSTER and LEAPER guide RNAs
2. Gene & protein sequences of:
    - AHI W725X
    - BMPR2 W298X
    - COL3A1 W1278X
    - FANCC W506X
    - IL2RG W237X
    - mIDUA wt
    - mIDUA W392X
    - MYBPC3 W1098X
    - PINK1 W437X
    - Dual-luciferase wt/wt
    - Dual-luciferase wt/amb

1. Cloning strategy for CLUSTER and LEAPER guide RNAs

Sequence of the multiple cloning site in our **gRNA cloning vector (pTS1033)** that is based on the psilencer 2.1 U6 hygro vector (ThermoFisher). The construct contains an R/G motif version 21 (ADAR recruiting domain). By using Hind-III and Bbs-I an antisense part (specificity domain plus a cluster of recruitment sequences) can be seamlessly added to the R/G motif. LEAPER guide RNAs can be introduced via Hind-III and BamHI. The guide RNAs are under control of an U6 promotor. A 6xT-stretch Pol3 termination signal should be added at the end of each oligonucleotide insert (LEAPER and CLUSTER).

```
1       GAGGGCCTAT TTCCCATGAT TCCTTCATAT TTGCATATAC GATACAAGGC TGTTAGAGAG
61      ATAATTAGAA TTAATTTGAC TGTAAACACA AAGATATTAG TACAAAATAC GTGACGTAGA
121     AAGTAATAAT TTCTTGGGTA GTTTGCAGTT TTAAAATTAT GTTTTAAAAT GGACTATCAT
181     ATGCTTACCG TAACTTGAAA GTATTTCGAT TTCTTGGGTT TATATATCTT GTGGAAAGGA
241     CGCGGGATCC GGTGTCGAGA AGAGGAGAAC AATATGCTAA ATGTTGTTCT CGTCTCCTCG
301     ACACCTAGTC TTCCGTAGTC GATAAGCTT
```

2. Gene & protein sequences:

Gene & protein sequence of the **AHI_W725X (pTS1159)** construct in the context of the pcDNA 3.1 vector, under control of the CMV promotor and the BGH polyA signal:

```
                10        20        30        40        50        60
1         GCTAGCATGCCTACAGCTGAGAGTGAAGCAAAAGTAAAAACCAAAGTTCGCTTTGAAGAA
1         Nhe-I  M   P   T   A   E   S   E   A   K   V   K   T   K   V   R   F   E   E

                70        80        90       100       110       120
61        TTGCTTAAGACCCACAGTGATCTAATGCGTGAAAAGAAAAAACTGAAGAAAAAACTTGTC
21         L   L   K   T   H   S   D   L   M   R   E   K   K   K   L   K   K   K   L   V

               130       140       150       160       170       180
121       AGGTCTGAAGAAAACATCTCACCTGACACTATTAGAAGCAATCTTCACTATATGAAAGAA
41         R   S   E   E   N   I   S   P   D   T   I   R   S   N   L   H   Y   M   K   E

               190       200       210       220       230       240
181       ACTACAAGTGATGATCCCGACACTATTAGAAGCAATCTTCCCCATATTAAAGAAACTACA
61         T   T   S   D   D   P   D   T   I   R   S   N   L   P   H   I   K   E   T   T

               250       260       270       280       290       300
241       AGTGATGATGTAAGTGCTGCTAACACTAACAACCTGAAGAAGAGCACGAGAGTCACTAAA
81         S   D   D   V   S   A   A   N   T   N   N   L   K   K   S   T   R   V   T   K

               310       320       330       340       350       360
301       AACAAATTGAGGAACACACAGTTAGCAACTGAAAATCCTAATGGTGATGCTAGTGTAGAG
101        N   K   L   R   N   T   Q   L   A   T   E   N   P   N   G   D   A   S   V   E

               370       380       390       400       410       420
361       GAAGACAAACAAGGAAAGCCAAATAAAAAGGTGATAAAGACGGTGCCCCAGTTGACTACA
121        E   D   K   Q   G   K   P   N   K   K   V   I   K   T   V   P   Q   L   T   T

               430       440       450       460       470       480
421       CAAGACCTGAAACCGGAAACTCCTGAGAATAAGGTTGATTCTACACACCAGAAAACACAT
141        Q   D   L   K   P   E   T   P   E   N   K   V   D   S   T   H   Q   K   T   H

               490       500       510       520       530       540
481       ACAAAGCCACAGCCAGGCGTTGATCATCAGAAAAGTGAGAAGGCAAATGAGGGAAGAGAA
161        T   K   P   Q   P   G   V   D   H   Q   K   S   E   K   A   N   E   G   R   E

               550       560       570       580       590       600
541       GAGACTGATTTAGAAGAGGATGAAGAATTGATGCAAGCATATCAGTGCCATGTAACTGAA
181        E   T   D   L   E   E   D   E   E   L   M   Q   A   Y   Q   C   H   V   T   E

               610       620       630       640       650       660
601       GAAATGGCAAAGGAGATTAAGAGGAAAATAAGAAAGAAACTGAAAGAACAGTTGACTTAC
201        E   M   A   K   E   I   K   R   K   I   R   K   K   L   K   E   Q   L   T   Y

               670       680       690       700       710       720
661       TTTCCCTCAGATACTTTATTCCATGATGACAAACTAAGCAGTGAAAAAAGGAAAAAGAAA
221        F   P   S   D   T   L   F   H   D   D   K   L   S   S   E   K   R   K   K   K

               730       740       750       760       770       780
721       AAGGAAGTTCCAGTCTTCTCTAAAGCTGAAACAAGTACATTGACCATCTCTGGTGACACA
241        K   E   V   P   V   F   S   K   A   E   T   S   T   L   T   I   S   G   D   T

               790       800       810       820       830       840
781       GTTGAAGGTGAACAAAAGAAAGAATCTTCAGTTAGATCAGTTTCTTCAGATTCTCATCAA
261        V   E   G   E   Q   K   K   E   S   S   V   R   S   V   S   S   D   S   H   Q
```

```
                850       860       870       880       890       900
841       GATGATGAAATAAGCTCAATGGAACAAAGCACAGAAGACAGCATGCAAGATGATACAAAA
281        D  D  E  I  S  S  M  E  Q  S  T  E  D  S  M  Q  D  D  T  K

                910       920       930       940       950       960
901       CCTAAACCAAAAAAAACAAAAAAGAAGACTAAAGCAGTTGCAGATAATAATGAAGATGTT
301        P  K  P  K  K  T  K  K  K  T  K  A  V  A  D  N  N  E  D  V

                970       980       990      1000      1010      1020
961       GATGGTGATGGTGTTCATGAAATAACAAGCCGAGATAGCCCGGTTTATCCCAAATGTTTG
321        D  G  D  G  V  H  E  I  T  S  R  D  S  P  V  Y  P  K  C  L

              1030      1040      1050      1060      1070      1080
1021      CTTGATGATGACCTTGTCTTGGGAGTTTACATTCACCGAACTGATAGACTTAAGTCAGAT
341        L  D  D  D  L  V  L  G  V  Y  I  H  R  T  D  R  L  K  S  D

              1090      1100      1110      1120      1130      1140
1081      TTTATGATTTCTCACCCAATGGTAAAAATTCATGTGGTTGATGAGCATACTGGTCAATAT
361        F  M  I  S  H  P  M  V  K  I  H  V  V  D  E  H  T  G  Q  Y

              1150      1160      1170      1180      1190      1200
1141      GTCAAGAAAGATGATAGTGGACGGCCTGTTTCATCTTACTATGAAAAAGAGAATGTGGAT
381        V  K  K  D  D  S  G  R  P  V  S  S  Y  Y  E  K  E  N  V  D

              1210      1220      1230      1240      1250      1260
1201      TATATTCTTCCTATTATGACCCAGCCATATGATTTTAAACAGTTAAAATCAAGACTTCCA
401        Y  I  L  P  I  M  T  Q  P  Y  D  F  K  Q  L  K  S  R  L  P

              1270      1280      1290      1300      1310      1320
1261      GAGTGGGAAGAACAAATTGTATTTAATGAAAATTTTCCCTATTTGCTTCGAGGCTCTGAT
421        E  W  E  E  Q  I  V  F  N  E  N  F  P  Y  L  L  R  G  S  D

              1330      1340      1350      1360      1370      1380
1321      GAGAGTCCTAAAGTCATCCTGTTCTTTGAGATTCTTGATTTCTTAAGCGTGGATGAAATT
441        E  S  P  K  V  I  L  F  F  E  I  L  D  F  L  S  V  D  E  I

              1390      1400      1410      1420      1430      1440
1381      AAGAATAATTCTGAGGTTCAAAACCAAGAATGTGGCTTTCGGAAAATTGCCTGGGCATTT
461        K  N  N  S  E  V  Q  N  Q  E  C  G  F  R  K  I  A  W  A  F

              1450      1460      1470      1480      1490      1500
1441      CTTAAGCTTCTGGGAGCCAATGGAAATGCAAACATCAACTCAAAACTTCGCTTGCAGCTA
481        L  K  L  L  G  A  N  G  N  A  N  I  N  S  K  L  R  L  Q  L

              1510      1520      1530      1540      1550      1560
1501      TATTACCCACCTACTAAGCCTCGATCCCCATTAAGTGTTGTTGAGGCATTTGAATGGTGG
501        Y  Y  P  P  T  K  P  R  S  P  L  S  V  V  E  A  F  E  W  W

              1570      1580      1590      1600      1610      1620
1561      TCAAAATGTCCAAGAAATCATTACCCATCAACACTGTACGTAACTGTAAGAGGACTGAAA
521        S  K  C  P  R  N  H  Y  P  S  T  L  Y  V  T  V  R  G  L  K

              1630      1640      1650      1660      1670      1680
1621      GTTCCAGACTGTATAAAGCCATCTTACCGCTCTATGATGGCTCTTCAGGAGGAAAAAGGT
541        V  P  D  C  I  K  P  S  Y  R  S  M  M  A  L  Q  E  E  K  G

              1690      1700      1710      1720      1730      1740
1681      AAACCAGTGCATTGTGAACGTCACCATGAGTCAAGCTCAGTAGACACAGAACCTGGATTA
561        K  P  V  H  C  E  R  H  H  E  S  S  S  V  D  T  E  P  G  L
```

```
                1750      1760      1770      1780      1790      1800
1741      GAAGAGTCAAAGGAAGTAATAAAGTGGAAACGACTCCCTGGGCAGGCTTGCCGTATCCCA
581        E  E  S  K  E  V  I  K  W  K  R  L  P  G  Q  A  C  R  I  P

                1810      1820      1830      1840      1850      1860
1801      AACAAACACCTCTTCTCACTAAATGCAGGAGAACGAGGATGTTTTTGTCTTGATTTCTCC
601        N  K  H  L  F  S  L  N  A  G  E  R  G  C  F  C  L  D  F  S

                1870      1880      1890      1900      1910      1920
1861      CACAATGGAAGAATATTAGCAGCAGCTTGTGCCAGCCGGGATGGATATCCAATTATTTTA
621        H  N  G  R  I  L  A  A  A  C  A  S  R  D  G  Y  P  I  I  L

                1930      1940      1950      1960      1970      1980
1921      TATGAAATTCCTTCTGGACGTTTCATGAGAGAATTGTGTGGCCACCTCAATATCATTTAT
641        Y  E  I  P  S  G  R  F  M  R  E  L  C  G  H  L  N  I  I  Y

                1990      2000      2010      2020      2030      2040
1981      GATCTTTCCTGGTCAAAAGATGATCACTACATCCTTACTTCATCATCTGATGGCACTGCC
661        D  L  S  W  S  K  D  D  H  Y  I  L  T  S  S  S  D  G  T  A

                2050      2060      2070      2080      2090      2100
2041      AGGATATGGAAAAATGAAATAAACAATACAAATACTTTCAGAGTTTTACCTCATCCTTCT
681        R  I  W  K  N  E  I  N  N  T  N  T  F  R  V  L  P  H  P  S

                2110      2120      2130      2140      2150      2160
2101      TTTGTTTACACGGCTAAATTCCATCCAGCTGTAAGAGAGCTAGTAGTTACAGGATGCTAT
701        F  V  Y  T  A  K  F  H  P  A  V  R  E  L  V  V  T  G  C  Y

                2170      2180      2190      2200      2210      2220
2161      GATTCCATGATACGGATATAGAAAGTTGAGATGAGAGAAGATTCTGCCATATTGGTCCGA
721        D  S  M  I  R  I  *  K  V  E  M  R  E  D  S  A  I  L  V  R

                2230      2240      2250      2260      2270      2280
2221      CAGTTTGACGTTCACAAAAGTTTTATCAACTCACTTTGTTTTGATACTGAAGGTCATCAT
741        Q  F  D  V  H  K  S  F  I  N  S  L  C  F  D  T  E  G  H  H

                2290      2300      2310      2320      2330      2340
2281      ATGTATTCAGGAGATTGTACAGGGGTGATTGTTGTTTGGAATACCTATGTCAAGATTAAT
761        M  Y  S  G  D  C  T  G  V  I  V  V  W  N  T  Y  V  K  I  N

                2350      2360      2370      2380      2390      2400
2341      GATTTGGAACATTCAGTGCACCACTGGACTATAAATAAGGAAATTAAAGAAACTGAGTTT
781        D  L  E  H  S  V  H  H  W  T  I  N  K  E  I  K  E  T  E  F

                2410      2420      2430      2440      2450      2460
2401      AAGGGAATTCCAATAAGTTATTTGGAGATTCATCCCAATGGAAAACGTTTGTTAATCCAT
801        K  G  I  P  I  S  Y  L  E  I  H  P  N  G  K  R  L  L  I  H

                2470      2480      2490      2500      2510      2520
2461      ACCAAAGACAGTACTTTGAGAATTATGGATCTCCGGATATTAGTAGCAAGGAAGTTTGTA
821        T  K  D  S  T  L  R  I  M  D  L  R  I  L  V  A  R  K  F  V

                2530      2540      2550      2560      2570      2580
2521      GGAGCAGCAAATTATCGGGAGAAGATTCATAGTACTTTGACTCCATGTGGGACTTTTCTG
841        G  A  A  N  Y  R  E  K  I  H  S  T  L  T  P  C  G  T  F  L

                2590      2600      2610      2620      2630      2640
2581      TTTGCTGGAAGTGAGGATGGTATAGTGTATGTTTGGAACCCAGAAACAGGAGAACAAGTA
861        F  A  G  S  E  D  G  I  V  Y  V  W  N  P  E  T  G  E  Q  V

                2650      2660      2670      2680      2690      2700
```

```
2641      GCCATGTATTCTGACTTGCCATTCAAGTCACCCATTCGAGACATTTCTTATCATCCATTT
881        A  M  Y  S  D  L  P  F  K  S  P  I  R  D  I  S  Y  H  P  F

               2710      2720      2730      2740      2750      2760
2701      GAAAATATGGTTGCATTCTGTGCATTTGGGCAAAATGAGCCAATTCTTCTGTATATTTAC
901        E  N  M  V  A  F  C  A  F  G  Q  N  E  P  I  L  L  Y  I  Y

               2770      2780      2790      2800      2810      2820
2761      GATTTCCATGTTGCCCAGCAGGAGGCTGAAATGTTCAAACGCTACAATGGAACATTTCCA
921        D  F  H  V  A  Q  Q  E  A  E  M  F  K  R  Y  N  G  T  F  P

               2830      2840      2850      2860      2870      2880
2821      TTACCTGGAATACACCAAAGTCAAGATGCCCTATGTACCTGTCCAAAACTACCCCATCAA
941        L  P  G  I  H  Q  S  Q  D  A  L  C  T  C  P  K  L  P  H  Q

               2890      2900      2910      2920      2930      2940
2881      GGCTCTTTTCAGATTGATGAATTTGTCCACACTGAAAGTTCTTCAACGAAGATGCAGCTA
961        G  S  F  Q  I  D  E  F  V  H  T  E  S  S  S  T  K  M  Q  L

               2950      2960      2970      2980      2990      3000
2941      GTAAAACAGAGGCTTGAAACTGTCACAGAGGTGATACGTTCCTGTGCTGCAAAAGTCAAC
981        V  K  Q  R  L  E  T  V  T  E  V  I  R  S  C  A  A  K  V  N

               3010      3020      3030      3040      3050      3060
3001      AAAAATCTCTCATTTACTTCACCACCAGCAGTTTCCTCACAACAGTCTAAGTTAAAGCAG
1001       K  N  L  S  F  T  S  P  P  A  V  S  S  Q  Q  S  K  L  K  Q

               3070      3080      3090      3100      3110      3120
3061      TCAAACATGCTGACCGCTCAAGAGATTCTACATCAGTTTGGTTTCACTCAGACCGGGATT
1021       S  N  M  L  T  A  Q  E  I  L  H  Q  F  G  F  T  Q  T  G  I

               3130      3140      3150      3160      3170      3180
3121      ATCAGCATAGAAAGAAAGCCTTGTAACCATCAGGTAGATACAGCACCAACGGTAGTGGCT
1041       I  S  I  E  R  K  P  C  N  H  Q  V  D  T  A  P  T  V  V  A

               3190      3200      3210      3220      3230      3240
3181      CTTTATGACTACACAGCGAATCGATCAGATGAACTAACCATCCATCGCGGAGACATTATC
1061       L  Y  D  Y  T  A  N  R  S  D  E  L  T  I  H  R  G  D  I  I

               3250      3260      3270      3280      3290      3300
3241      CGAGTGTTTTTCAAAGATAATGAAGACTGGTGGTATGGCAGCATAGGAAAGGGACAGGAA
1081       R  V  F  F  K  D  N  E  D  W  W  Y  G  S  I  G  K  G  Q  E

               3310      3320      3330      3340      3350      3360
3301      GGTTATTTTCCAGCTAATCATGTGGCTAGTGAAACACTGTATCAAGAACTGCCTCCTGAG
1101       G  Y  F  P  A  N  H  V  A  S  E  T  L  Y  Q  E  L  P  P  E

               3370      3380      3390      3400      3410      3420
3361      ATAAAGGAGCGATCCCCTCCTTTAAGCCCTGAGGAAAAAACTAAAATAGAAAAATCTCCA
1121       I  K  E  R  S  P  P  L  S  P  E  E  K  T  K  I  E  K  S  P

               3430      3440      3450      3460      3470      3480
3421      GCTCCTCAAAAGCAATCAATCAATAAGAACAAGTCCCAGGACTTCAGACTAGGCTCAGAA
1141       A  P  Q  K  Q  S  I  N  K  N  K  S  Q  D  F  R  L  G  S  E

               3490      3500      3510      3520      3530      3540
3481      TCTATGACACATTCTGAAATGAGAAAAGAACAGAGCCATGAGGACCAAGGACACATAATG
1161       S  M  T  H  S  E  M  R  K  E  Q  S  H  E  D  Q  G  H  I  M

               3550      3560      3570      3580      3590      3600
3541      GATACACGGATGAGGAAGAACAAGCAAGCAGGCAGAAAAGTCACTCTAATAGAGTAGCGG
```

```
1181      D  T  R  M  R  K  N  K  Q  A  G  R  K  V  T  L  I  E  *

3601      CCGCTCGAG
1201        Xho-I
```

Gene & protein sequence of the **BMPR2_W298X (pTS1160)** construct in the context of the pcDNA 3.1 vector, under control of the CMV promotor and the BGH polyA signal:

```
                10        20        30        40        50        60
1       GCTAGCATGACTTCCTCGCTGCAGCGGCCCTGGCGGGTGCCCTGGCTACCATGGACCATC
1       Nhe-I  M  T  S  S  L  Q  R  P  W  R  V  P  W  L  P  W  T  I

                70        80        90       100       110       120
61      CTGCTGGTCAGCGCTGCGGCTGCTTCGCAGAATCAAGAACGGCTATGTGCGTTTAAAGAT
21       L  L  V  S  A  A  A  A  S  Q  N  Q  E  R  L  C  A  F  K  D

               130       140       150       160       170       180
121     CCGTATCAGCAAGACCTTGGGATAGGTGAGAGTAGAATCTCTCATGAAAATGGGACAATA
41       P  Y  Q  Q  D  L  G  I  G  E  S  R  I  S  H  E  N  G  T  I

               190       200       210       220       230       240
181     TTATGCTCGAAAGGTAGCACCTGCTATGGCCTTTGGGAGAAATCAAAAGGGGACATAAAT
61       L  C  S  K  G  S  T  C  Y  G  L  W  E  K  S  K  G  D  I  N

               250       260       270       280       290       300
241     CTTGTAAAACAAGGATGTTGGTCTCACATTGGAGATCCCCAAGAGTGTCACTATGAAGAA
81       L  V  K  Q  G  C  W  S  H  I  G  D  P  Q  E  C  H  Y  E  E

               310       320       330       340       350       360
301     TGTGTAGTAACTACCACTCCTCCCTCAATTCAGAATGGAACATACCGTTTCTGCTGTTGT
101      C  V  V  T  T  T  P  P  S  I  Q  N  G  T  Y  R  F  C  C  C

               370       380       390       400       410       420
361     AGCACAGATTTATGTAATGTCAACTTTACTGAGAATTTTCCACCTCCTGACACAACACCA
121      S  T  D  L  C  N  V  N  F  T  E  N  F  P  P  P  D  T  T  P

               430       440       450       460       470       480
421     CTCAGTCCACCTCATTCATTTAACCGAGATGAGACAATAATCATTGCTTTGGCATCAGTC
141      L  S  P  P  H  S  F  N  R  D  E  T  I  I  I  A  L  A  S  V

               490       500       510       520       530       540
481     TCTGTATTAGCTGTTTTGATAGTTGCCTTATGCTTTGGATACAGAATGTTGACAGGAGAC
161      S  V  L  A  V  L  I  V  A  L  C  F  G  Y  R  M  L  T  G  D

               550       560       570       580       590       600
541     CGTAAACAAGGTCTTCACAGTATGAACATGATGGAGGCAGCAGCATCCGAACCCTCTCTT
181      R  K  Q  G  L  H  S  M  N  M  M  E  A  A  A  S  E  P  S  L

               610       620       630       640       650       660
601     GATCTAGATAATCTGAAACTGTTGGAGCTGATTGGCCGAGGTCGATATGGAGCAGTATAT
201      D  L  D  N  L  K  L  L  E  L  I  G  R  G  R  Y  G  A  V  Y

               670       680       690       700       710       720
661     AAAGGCTCCTTGGATGAGCGTCCAGTTGCTGTAAAAGTGTTTTCCTTTGCAAACCGTCAG
221      K  G  S  L  D  E  R  P  V  A  V  K  V  F  S  F  A  N  R  Q

               730       740       750       760       770       780
721     AATTTTATCAACGAAAAGAACATTTACAGAGTGCCTTTGATGGAACATGACAACATTGCC
241      N  F  I  N  E  K  N  I  Y  R  V  P  L  M  E  H  D  N  I  A

               790       800       810       820       830       840
781     CGCTTTATAGTTGGAGATGAGAGAGTCACTGCAGATGGACGCATGGAATATTTGCTTGTG
261      R  F  I  V  G  D  E  R  V  T  A  D  G  R  M  E  Y  L  L  V

               850       860       870       880       890       900
```

```
841       ATGGAGTACTATCCCAATGGATCTTTATGCAAGTATTTAAGTCTCCACACAAGTGACTAG
281        M  E  Y  Y  P  N  G  S  L  C  K  Y  L  S  L  H  T  S  D  *

                 910       920       930       940       950       960
901       GTAAGCTCTTGCCGTCTTGCTCATTCTGTTACTAGAGGACTGGCTTATCTTCACACAGAA
301        V  S  S  C  R  L  A  H  S  V  T  R  G  L  A  Y  L  H  T  E

                 970       980       990      1000      1010      1020
961       TTACCACGAGGAGATCATTATAAACCTGCAATTTCCCATCGAGATTTAAACAGCAGAAAT
321        L  P  R  G  D  H  Y  K  P  A  I  S  H  R  D  L  N  S  R  N

               1030      1040      1050      1060      1070      1080
1021      GTCCTAGTGAAAAATGATGGAACCTGTGTTATTAGTGACTTTGGACTGTCCATGAGGCTG
341        V  L  V  K  N  D  G  T  C  V  I  S  D  F  G  L  S  M  R  L

               1090      1100      1110      1120      1130      1140
1081      ACTGGAAATAGACTGGTGCGCCCAGGGGAGGAAGATAATGCAGCCATAAGCGAGGTTGGC
361        T  G  N  R  L  V  R  P  G  E  E  D  N  A  A  I  S  E  V  G

               1150      1160      1170      1180      1190      1200
1141      ACTATCAGATATATGGCACCAGAAGTGCTAGAAGGAGCTGTGAACTTGAGGGACTGTGAA
381        T  I  R  Y  M  A  P  E  V  L  E  G  A  V  N  L  R  D  C  E

               1210      1220      1230      1240      1250      1260
1201      TCAGCTTTGAAACAAGTAGACATGTATGCTCTTGGACTAATCTATTGGGAGATATTTATG
401        S  A  L  K  Q  V  D  M  Y  A  L  G  L  I  Y  W  E  I  F  M

               1270      1280      1290      1300      1310      1320
1261      AGATGTACAGACCTCTTCCCAGGGGAATCCGTACCAGAGTACCAGATGGCTTTTCAGACA
421        R  C  T  D  L  F  P  G  E  S  V  P  E  Y  Q  M  A  F  Q  T

               1330      1340      1350      1360      1370      1380
1321      GAGGTTGGAAACCATCCCACTTTTGAGGATATGCAGGTTCTCGTGTCTAGGGAAAAACAG
441        E  V  G  N  H  P  T  F  E  D  M  Q  V  L  V  S  R  E  K  Q

               1390      1400      1410      1420      1430      1440
1381      AGACCCAAGTTCCCAGAAGCCTGGAAAGAAAATAGCCTGGCAGTGAGGTCACTCAAGGAG
461        R  P  K  F  P  E  A  W  K  E  N  S  L  A  V  R  S  L  K  E

               1450      1460      1470      1480      1490      1500
1441      ACAATCGAAGACTGTTGGGACCAGGATGCAGAGGCTCGGCTTACTGCACAGTGTGCTGAG
481        T  I  E  D  C  W  D  Q  D  A  E  A  R  L  T  A  Q  C  A  E

               1510      1520      1530      1540      1550      1560
1501      GAAAGGATGGCTGAACTTATGATGATTTGGGAAAGAAACAAATCTGTGAGCCCAACAGTC
501        E  R  M  A  E  L  M  M  I  W  E  R  N  K  S  V  S  P  T  V

               1570      1580      1590      1600      1610      1620
1561      AATCCAATGTCTACTGCTATGCAGAATGAACGCAACCTGTCACATAATAGGCGTGTGCCA
521        N  P  M  S  T  A  M  Q  N  E  R  N  L  S  H  N  R  R  V  P

               1630      1640      1650      1660      1670      1680
1621      AAAATTGGTCCTTATCCAGATTATTCTTCCTCCTCATACATTGAAGACTCTATCCATCAT
541        K  I  G  P  Y  P  D  Y  S  S  S  S  Y  I  E  D  S  I  H  H

               1690      1700      1710      1720      1730      1740
1681      ACTGACAGCATCGTGAAGAATATTTCCTCTGAGCATTCTATGTCCAGCACACCTTTGACT
561        T  D  S  I  V  K  N  I  S  S  E  H  S  M  S  S  T  P  L  T

               1750      1760      1770      1780      1790      1800
1741      ATAGGGGAAAAAAAACCGAAATTCAATTAACTATGAACGACAGCAAGCACAAGCTCGAATC
```

```
581        I  G  E  K  N  R  N  S  I  N  Y  E  R  Q  Q  A  Q  A  R  I

                1810       1820       1830       1840       1850       1860
1801       CCCAGCCCTGAAACAAGTGTCACCAGCCTCTCCACCAACACAACAACCACAAACACCACA
601         P  S  P  E  T  S  V  T  S  L  S  T  N  T  T  T  T  N  T  T

                1870       1880       1890       1900       1910       1920
1861       GGACTCACGCCAAGTACTGGCATGACTACTATATCTGAGATGCCATACCCAGATGAAACA
621         G  L  T  P  S  T  G  M  T  T  I  S  E  M  P  Y  P  D  E  T

                1930       1940       1950       1960       1970       1980
1921       AATCTGCATACCACAAATGTTGCACAGTCAATTGGGCCAACCCCTGTCTGCTTACAGCTG
641         N  L  H  T  T  N  V  A  Q  S  I  G  P  T  P  V  C  L  Q  L

                1990       2000       2010       2020       2030       2040
1981       ACAGAAGAAGACTTGGAAACCAACAAGCTAGACCCAAAAGAAGTTGATAAGAACCTCAAG
661         T  E  E  D  L  E  T  N  K  L  D  P  K  E  V  D  K  N  L  K

                2050       2060       2070       2080       2090       2100
2041       GAAAGCTCTGATGAGAATCTCATGGAGCACTCTCTTAAACAGTTCAGTGGCCCAGACCCA
681         E  S  S  D  E  N  L  M  E  H  S  L  K  Q  F  S  G  P  D  P

                2110       2120       2130       2140       2150       2160
2101       CTGAGCAGTACTAGTTCTAGCTTGCTTTACCCACTCATAAAACTTGCAGTAGAAGCAACT
701         L  S  S  T  S  S  S  L  L  Y  P  L  I  K  L  A  V  E  A  T

                2170       2180       2190       2200       2210       2220
2161       GGACAGCAGGACTTCACACAGACTGCAAATGGCCAAGCATGTTTGATTCCTGATGTTCTG
721         G  Q  Q  D  F  T  Q  T  A  N  G  Q  A  C  L  I  P  D  V  L

                2230       2240       2250       2260       2270       2280
2221       CCTACTCAGATCTATCCTCTCCCCAAGCAGCAGAACCTTCCCAAGAGACCTACTAGTTTG
741         P  T  Q  I  Y  P  L  P  K  Q  Q  N  L  P  K  R  P  T  S  L

                2290       2300       2310       2320       2330       2340
2281       CCTTTGAACACCAAAAATTCAACAAAAGAGCCCCGGCTAAAATTTGGCAGCAAGCACAAA
761         P  L  N  T  K  N  S  T  K  E  P  R  L  K  F  G  S  K  H  K

                2350       2360       2370       2380       2390       2400
2341       TCAAACTTGAAACAAGTCGAAACTGGAGTTGCCAAGATGAATACAATCAATGCAGCAGAA
781         S  N  L  K  Q  V  E  T  G  V  A  K  M  N  T  I  N  A  A  E

                2410       2420       2430       2440       2450       2460
2401       CCTCATGTGGTGACAGTCACCATGAATGGTGTGGCAGGTAGAAACCACAGTGTTAACTCC
801         P  H  V  V  T  V  T  M  N  G  V  A  G  R  N  H  S  V  N  S

                2470       2480       2490       2500       2510       2520
2461       CATGCTGCCACAACCCAATATGCCAATGGGACAGTACTATCTGGCCAAACAACCAACATA
821         H  A  A  T  T  Q  Y  A  N  G  T  V  L  S  G  Q  T  T  N  I

                2530       2540       2550       2560       2570       2580
2521       GTGACACATAGGGCCCAAGAAATGTTGCAGAATCAGTTTATTGGTGAGGACACCCGGCTG
841         V  T  H  R  A  Q  E  M  L  Q  N  Q  F  I  G  E  D  T  R  L

                2590       2600       2610       2620       2630       2640
2581       AATATTAATTCCAGTCCTGATGAGCATGAGCCTTTACTGAGACGAGAGCAACAAGCTGGC
861         N  I  N  S  S  P  D  E  H  E  P  L  L  R  R  E  Q  Q  A  G

                2650       2660       2670       2680       2690       2700
2641       CATGATGAAGGTGTTCTGGATCGTCTTGTGGACAGGAGGGAACGGCCACTAGAAGGTGGC
881         H  D  E  G  V  L  D  R  L  V  D  R  R  E  R  P  L  E  G  G
```

```
                2710      2720      2730      2740      2750      2760
2701       CGAACTAATTCCAATAACAACAACAGCAATCCATGTTCAGAACAAGATGTTCTTGCACAG
901         R  T  N  S  N  N  N  N  S  N  P  C  S  E  Q  D  V  L  A  Q

                2770      2780      2790      2800      2810      2820
2761       GGTGTTCCAAGCACAGCAGCAGATCCTGGGCCATCAAAGCCCAGAAGAGCACAGAGGCCT
921         G  V  P  S  T  A  A  D  P  G  P  S  K  P  R  R  A  Q  R  P

                2830      2840      2850      2860      2870      2880
2821       AATTCTCTGGATCTTTCAGCCACAAATGTCCTGGATGGCAGCAGTATACAGATAGGTGAG
941         N  S  L  D  L  S  A  T  N  V  L  D  G  S  S  I  Q  I  G  E

                2890      2900      2910      2920      2930      2940
2881       TCAACACAAGATGGCAAATCAGGATCAGGTGAAAAGATCAAGAAACGTGTGAAAACTCCC
961         S  T  Q  D  G  K  S  G  S  G  E  K  I  K  K  R  V  K  T  P

                2950      2960      2970      2980      2990      3000
2941       TATTCTCTTAAGCGGTGGCGCCCCTCCACCTGGGTCATCTCCACTGAATCGCTGGACTGT
981         Y  S  L  K  R  W  R  P  S  T  W  V  I  S  T  E  S  L  D  C

                3010      3020      3030      3040      3050      3060
3001       GAAGTCAACAATAATGGCAGTAACAGGGCAGTTCATTCCAAATCCAGCACTGCTGTTTAC
1001        E  V  N  N  N  G  S  N  R  A  V  H  S  K  S  S  T  A  V  Y

                3070      3080      3090      3100      3110      3120
3061       CTTGCAGAAGGAGGCACTGCTACAACCATGGTGTCTAAAGATATAGGAATGAACTGTCTG
1021        L  A  E  G  G  T  A  T  T  M  V  S  K  D  I  G  M  N  C  L

                3130
3121       TGAGCGGCCGCTCGAG
1041        *        Xho-I
```

Gene & protein sequence of the **COL3A1_W1278X (pTS1161)** construct in the context of the pcDNA 3.1 vector, under control of the CMV promotor and the BGH polyA signal:

```
                10        20        30        40        50        60
1       GCTAGCATGATGAGCTTTGTGCAAAAGGGGAGCTGGCTACTTCTCGCTCTGCTTCATCCC
1       Nhe-I  M  M  S  F  V  Q  K  G  S  W  L  L  L  A  L  L  H  P

                70        80        90       100       110       120
61      ACTATTATTTTGGCACAACAGGAAGCTGTTGAAGGAGGATGTTCCCATCTTGGTCAGTCC
21       T  I  I  L  A  Q  Q  E  A  V  E  G  G  C  S  H  L  G  Q  S

               130       140       150       160       170       180
121     TATGCGGATAGAGATGTCTGGAAGCCAGAACCATGCCAAATATGTGTCTGTGACTCAGGA
41       Y  A  D  R  D  V  W  K  P  E  P  C  Q  I  C  V  C  D  S  G

               190       200       210       220       230       240
181     TCCGTTCTCTGCGATGACATAATATGTGACGATCAAGAATTAGACTGCCCCAACCCAGAA
61       S  V  L  C  D  D  I  I  C  D  D  Q  E  L  D  C  P  N  P  E

               250       260       270       280       290       300
241     ATTCCATTTGGAGAATGTTGTGCAGTTTGCCCACAGCCTCCAACTGCTCCTACTCGCCCT
81       I  P  F  G  E  C  C  A  V  C  P  Q  P  P  T  A  P  T  R  P

               310       320       330       340       350       360
301     CCTAATGGTCAAGGACCTCAAGGCCCCAAGGGAGATCCAGGCCCTCCTGGTATTCCTGGG
101      P  N  G  Q  G  P  Q  G  P  K  G  D  P  G  P  P  G  I  P  G

               370       380       390       400       410       420
361     AGAAATGGTGACCCTGGTATTCCAGGACAACCAGGGTCCCCTGGTTCTCCTGGCCCCCCT
121      R  N  G  D  P  G  I  P  G  Q  P  G  S  P  G  S  P  G  P  P

               430       440       450       460       470       480
421     GGAATCTGTGAATCATGCCCTACTGGTCCTCAGAACTATTCTCCCCAGTATGATTCATAT
141      G  I  C  E  S  C  P  T  G  P  Q  N  Y  S  P  Q  Y  D  S  Y

               490       500       510       520       530       540
481     GATGTCAAGTCTGGAGTAGCAGTAGGAGGACTCGCAGGCTATCCTGGACCAGCTGGCCCC
161      D  V  K  S  G  V  A  V  G  G  L  A  G  Y  P  G  P  A  G  P

               550       560       570       580       590       600
541     CCAGGCCCTCCCGGTCCCCCTGGTACATCTGGTCATCCTGGTTCCCCTGGATCTCCAGGA
181      P  G  P  P  G  P  P  G  T  S  G  H  P  G  S  P  G  S  P  G

               610       620       630       640       650       660
601     TACCAAGGACCCCCTGGTGAACCTGGGCAAGCTGGTCCTTCAGGCCCTCCAGGACCTCCT
201      Y  Q  G  P  P  G  E  P  G  Q  A  G  P  S  G  P  P  G  P  P

               670       680       690       700       710       720
661     GGTGCTATAGGTCCATCTGGTCCTGCTGGAAAAGATGGAGAATCAGGTAGACCCGGACGA
221      G  A  I  G  P  S  G  P  A  G  K  D  G  E  S  G  R  P  G  R

               730       740       750       760       770       780
721     CCTGGAGAGCGAGGATTGCCTGGACCTCCAGGTATCAAAGGTCCAGCTGGGATACCTGGA
241      P  G  E  R  G  L  P  G  P  P  G  I  K  G  P  A  G  I  P  G

               790       800       810       820       830       840
781     TTCCCTGGTATGAAAGGACACAGAGGCTTCGATGGACGAAATGGAGAAAAGGGTGAAACA
261      F  P  G  M  K  G  H  R  G  F  D  G  R  N  G  E  K  G  E  T

               850       860       870       880       890       900
```

```
841       GGTGCTCCTGGATTAAAGGGTGAAAATGGTCTTCCAGGCGAAAATGGAGCTCCTGGACCC
281        G  A  P  G  L  K  G  E  N  G  L  P  G  E  N  G  A  P  G  P

                 910       920       930       940       950       960
901       ATGGGTCCAAGAGGGGCTCCTGGTGAGCGAGGACGGCCAGGACTTCCTGGGGCTGCAGGT
301        M  G  P  R  G  A  P  G  E  R  G  R  P  G  L  P  G  A  A  G

                 970       980       990      1000      1010      1020
961       GCTCGGGGTAATGACGGTGCTCGAGGCAGTGATGGTCAACCAGGCCCTCCTGGTCCTCCT
321        A  R  G  N  D  G  A  R  G  S  D  G  Q  P  G  P  P  G  P  P

               1030      1040      1050      1060      1070      1080
1021      GGAACTGCCGGATTCCCTGGATCCCCTGGTGCTAAGGGTGAAGTTGGACCTGCAGGGTCT
341        G  T  A  G  F  P  G  S  P  G  A  K  G  E  V  G  P  A  G  S

               1090      1100      1110      1120      1130      1140
1081      CCTGGTTCAAATGGTGCCCCTGGACAAAGAGGAGAACCTGGACCTCAGGGACACGCTGGT
361        P  G  S  N  G  A  P  G  Q  R  G  E  P  G  P  Q  G  H  A  G

               1150      1160      1170      1180      1190      1200
1141      GCTCAAGGTCCTCCTGGCCCTCCTGGGATTAATGGTAGTCCTGGTGGTAAAGGCGAAATG
381        A  Q  G  P  P  G  P  P  G  I  N  G  S  P  G  G  K  G  E  M

               1210      1220      1230      1240      1250      1260
1201      GGTCCCGCTGGCATTCCTGGAGCTCCTGGACTGATGGGAGCCCGGGGTCCTCCAGGACCA
401        G  P  A  G  I  P  G  A  P  G  L  M  G  A  R  G  P  P  G  P

               1270      1280      1290      1300      1310      1320
1261      GCCGGTGCTAATGGTGCTCCTGGACTGCGAGGTGGTGCAGGTGAGCCTGGTAAGAATGGT
421        A  G  A  N  G  A  P  G  L  R  G  G  A  G  E  P  G  K  N  G

               1330      1340      1350      1360      1370      1380
1321      GCCAAAGGAGAGCCCGGACCACGTGGTGAACGCGGTGAGGCTGGTATTCCAGGTGTTCCA
441        A  K  G  E  P  G  P  R  G  E  R  G  E  A  G  I  P  G  V  P

               1390      1400      1410      1420      1430      1440
1381      GGAGCTAAAGGCGAAGATGGCAAGGATGGATCACCTGGAGAACCTGGTGCAAATGGGCTT
461        G  A  K  G  E  D  G  K  D  G  S  P  G  E  P  G  A  N  G  L

               1450      1460      1470      1480      1490      1500
1441      CCAGGAGCTGCAGGAGAAAGGGGTGCCCCTGGGTTCCGAGGACCTGCTGGACCAAATGGC
481        P  G  A  A  G  E  R  G  A  P  G  F  R  G  P  A  G  P  N  G

               1510      1520      1530      1540      1550      1560
1501      ATCCCAGGAGAAAAGGGTCCTGCTGGAGAGCGTGGTGCTCCAGGCCCTGCAGGGCCCAGA
501        I  P  G  E  K  G  P  A  G  E  R  G  A  P  G  P  A  G  P  R

               1570      1580      1590      1600      1610      1620
1561      GGAGCTGCTGGAGAACCTGGCAGAGATGGCGTCCCTGGAGGTCCAGGAATGAGGGGCATG
521        G  A  A  G  E  P  G  R  D  G  V  P  G  G  P  G  M  R  G  M

               1630      1640      1650      1660      1670      1680
1621      CCCGGAAGTCCAGGAGGACCAGGAAGTGATGGGAAACCAGGGCCTCCCGGAAGTCAAGGA
541        P  G  S  P  G  G  P  G  S  D  G  K  P  G  P  P  G  S  Q  G

               1690      1700      1710      1720      1730      1740
1681      GAAAGTGGTCGACCAGGTCCTCCTGGGCCATCTGGTCCCCGAGGTCAGCCTGGTGTCATG
561        E  S  G  R  P  G  P  P  G  P  S  G  P  R  G  Q  P  G  V  M

               1750      1760      1770      1780      1790      1800
1741      GGCTTCCCCGGTCCTAAAGGAAATGATGGTGCTCCTGGTAAGAATGGAGAACGAGGTGGC
```

```
581        G  F  P  G  P  K  G  N  D  G  A  P  G  K  N  G  E  R  G  G

                1810       1820       1830       1840       1850       1860
1801       CCTGGAGGACCTGGCCCTCAGGGTCCTCCTGGAAAGAATGGTGAAACTGGACCTCAGGGA
601        P  G  G  P  G  P  Q  G  P  P  G  K  N  G  E  T  G  P  Q  G

                1870       1880       1890       1900       1910       1920
1861       CCCCCAGGGCCTACTGGGCCTGGTGGTGACAAAGGAGACACAGGACCCCCTGGTCCACAA
621        P  P  G  P  T  G  P  G  G  D  K  G  D  T  G  P  P  G  P  Q

                1930       1940       1950       1960       1970       1980
1921       GGATTACAAGGCTTGCCTGGTACAGGTGGTCCTCCAGGAGAAAATGGAAAACCTGGGGAA
641        G  L  Q  G  L  P  G  T  G  G  P  P  G  E  N  G  K  P  G  E

                1990       2000       2010       2020       2030       2040
1981       CCAGGTCCAAAGGGTGATGCCGGTGCACCTGGAGCTCCAGGAGGCAAGGGTGATGCTGGT
661        P  G  P  K  G  D  A  G  A  P  G  A  P  G  G  K  G  D  A  G

                2050       2060       2070       2080       2090       2100
2041       GCCCCTGGTGAACGTGGACCTCCTGGATTGGCAGGGGCCCCAGGACTTAGAGGTGGAGCT
681        A  P  G  E  R  G  P  P  G  L  A  G  A  P  G  L  R  G  G  A

                2110       2120       2130       2140       2150       2160
2101       GGTCCCCCTGGTCCCGAAGGAGGAAAGGGTGCTGCTGGTCCTCCTGGGCCACCTGGTGCT
701        G  P  P  G  P  E  G  G  K  G  A  A  G  P  P  G  P  P  G  A

                2170       2180       2190       2200       2210       2220
2161       GCTGGTACTCCTGGTCTGCAAGGAATGCCTGGAGAAAGAGGAGGTCTTGGAAGTCCTGGT
721        A  G  T  P  G  L  Q  G  M  P  G  E  R  G  G  L  G  S  P  G

                2230       2240       2250       2260       2270       2280
2221       CCAAAGGGTGACAAGGGTGAACCAGGCGGTCCAGGTGCTGATGGTGTCCCAGGGAAAGAT
741        P  K  G  D  K  G  E  P  G  G  P  G  A  D  G  V  P  G  K  D

                2290       2300       2310       2320       2330       2340
2281       GGCCCAAGGGGTCCTACTGGTCCTATTGGTCCTCCTGGCCCAGCTGGCCAGCCTGGAGAT
761        G  P  R  G  P  T  G  P  I  G  P  P  G  P  A  G  Q  P  G  D

                2350       2360       2370       2380       2390       2400
2341       AAGGGTGAAGGTGGTGCCCCCGGACTTCCAGGTATAGCTGGACCTCGTGGTAGCCCTGGT
781        K  G  E  G  G  A  P  G  L  P  G  I  A  G  P  R  G  S  P  G

                2410       2420       2430       2440       2450       2460
2401       GAGAGAGGTGAAACTGGCCCTCCAGGACCTGCTGGTTTCCCTGGTGCTCCTGGACAGAAT
801        E  R  G  E  T  G  P  P  G  P  A  G  F  P  G  A  P  G  Q  N

                2470       2480       2490       2500       2510       2520
2461       GGTGAACCTGGTGGTAAAGGAGAAAGAGGGGCTCCGGGTGAGAAAGGTGAAGGAGGCCCT
821        G  E  P  G  G  K  G  E  R  G  A  P  G  E  K  G  E  G  G  P

                2530       2540       2550       2560       2570       2580
2521       CCTGGAGTTGCAGGACCCCCTGGAGGTTCTGGACCTGCTGGTCCTCCTGGTCCCCAAGGT
841        P  G  V  A  G  P  P  G  G  S  G  P  A  G  P  P  G  P  Q  G

                2590       2600       2610       2620       2630       2640
2581       GTCAAAGGTGAACGTGGCAGTCCTGGTGGACCTGGTGCTGCTGGCTTCCCTGGTGCTCGT
861        V  K  G  E  R  G  S  P  G  G  P  G  A  A  G  F  P  G  A  R

                2650       2660       2670       2680       2690       2700
2641       GGTCTTCCTGGTCCTCCTGGTAGTAATGGTAACCCAGGACCCCCAGGTCCCAGCGGTTCT
881        G  L  P  G  P  P  G  S  N  G  N  P  G  P  P  G  P  S  G  S
```

```
              2710      2720      2730      2740      2750      2760
2701      CCAGGCAAGGATGGGCCCCCAGGTCCTGCGGGTAACACTGGTGCTCCTGGCAGCCCTGGA
901        P  G  K  D  G  P  P  G  P  A  G  N  T  G  A  P  G  S  P  G

              2770      2780      2790      2800      2810      2820
2761      GTGTCTGGACCAAAAGGTGATGCTGGCCAACCAGGAGAGAAGGGATCGCCTGGTGCCCAG
921        V  S  G  P  K  G  D  A  G  Q  P  G  E  K  G  S  P  G  A  Q

              2830      2840      2850      2860      2870      2880
2821      GGCCCACCAGGAGCTCCAGGCCCACTTGGGATTGCTGGGATCACTGGAGCACGGGGTCTT
941        G  P  P  G  A  P  G  P  L  G  I  A  G  I  T  G  A  R  G  L

              2890      2900      2910      2920      2930      2940
2881      GCAGGACCACCAGGCATGCCAGGTCCTAGGGGAAGCCCTGGCCCTCAGGGTGTCAAGGGT
961        A  G  P  P  G  M  P  G  P  R  G  S  P  G  P  Q  G  V  K  G

              2950      2960      2970      2980      2990      3000
2941      GAAAGTGGGAAACCAGGAGCTAACGGTCTCAGTGGAGAACGTGGTCCCCCTGGACCCCAG
981        E  S  G  K  P  G  A  N  G  L  S  G  E  R  G  P  P  G  P  Q

              3010      3020      3030      3040      3050      3060
3001      GGTCTTCCTGGTCTGGCTGGTACAGCTGGTGAACCTGGAAGAGATGGAAACCCTGGATCA
1001       G  L  P  G  L  A  G  T  A  G  E  P  G  R  D  G  N  P  G  S

              3070      3080      3090      3100      3110      3120
3061      GATGGTCTTCCAGGCCGAGATGGATCTCCTGGTGGCAAGGGTGATCGTGGTGAAAATGGC
1021       D  G  L  P  G  R  D  G  S  P  G  G  K  G  D  R  G  E  N  G

              3130      3140      3150      3160      3170      3180
3121      TCTCCTGGTGCCCCTGGCGCTCCTGGTCATCCAGGCCCACCTGGTCCTGTCGGTCCAGCT
1041       S  P  G  A  P  G  A  P  G  H  P  G  P  P  G  P  V  G  P  A

              3190      3200      3210      3220      3230      3240
3181      GGAAAGAGTGGTGACAGAGGAGAAAGTGGCCCTGCTGGCCCTGCTGGTGCTCCCGGTCCT
1061       G  K  S  G  D  R  G  E  S  G  P  A  G  P  A  G  A  P  G  P

              3250      3260      3270      3280      3290      3300
3241      GCTGGTTCCCGAGGTGCTCCTGGTCCTCAAGGCCCACGTGGTGACAAAGGTGAAACAGGT
1081       A  G  S  R  G  A  P  G  P  Q  G  P  R  G  D  K  G  E  T  G

              3310      3320      3330      3340      3350      3360
3301      GAACGTGGAGCTGCTGGCATCAAAGGACATCGAGGATTCCCTGGTAATCCAGGTGCCCCA
1101       E  R  G  A  A  G  I  K  G  H  R  G  F  P  G  N  P  G  A  P

              3370      3380      3390      3400      3410      3420
3361      GGTTCTCCAGGCCCTGCTGGTCAGCAGGGTGCAATCGGCAGTCCAGGACCTGCAGGCCCC
1121       G  S  P  G  P  A  G  Q  Q  G  A  I  G  S  P  G  P  A  G  P

              3430      3440      3450      3460      3470      3480
3421      AGAGGACCTGTTGGACCCAGTGGACCTCCTGGCAAAGATGGAACCAGTGGACATCCAGGT
1141       R  G  P  V  G  P  S  G  P  P  G  K  D  G  T  S  G  H  P  G

              3490      3500      3510      3520      3530      3540
3481      CCCATTGGACCACCAGGGCCTCGAGGTAACAGAGGTGAAAGAGGATCTGAGGGCTCCCCA
1161       P  I  G  P  P  G  P  R  G  N  R  G  E  R  G  S  E  G  S  P

              3550      3560      3570      3580      3590      3600
3541      GGCCACCCAGGGCAACCAGGCCCTCCTGGACCTCCTGGTGCCCCTGGTCCTTGCTGTGGT
1181       G  H  P  G  Q  P  G  P  P  G  P  P  G  A  P  G  P  C  C  G
```

```
             3610      3620      3630      3640      3650      3660
3601   GGTGTTGGAGCCGCTGCCATTGCTGGGATTGGAGGTGAAAAAGCTGGCGGTTTTGCCCCG
1201    G  V  G  A  A  A  I  A  G  I  G  G  E  K  A  G  G  F  A  P

             3670      3680      3690      3700      3710      3720
3661   TATTATGGAGATGAACCAATGGATTTCAAAATCAACACCGATGAGATTATGACTTCACTC
1221    Y  Y  G  D  E  P  M  D  F  K  I  N  T  D  E  I  M  T  S  L

             3730      3740      3750      3760      3770      3780
3721   AAGTCTGTTAATGGACAAATAGAAAGCCTCATTAGTCCTGATGGTTCTCGTAAAAACCCC
1241    K  S  V  N  G  Q  I  E  S  L  I  S  P  D  G  S  R  K  N  P

             3790      3800      3810      3820      3830      3840
3781   GCTAGAAACTGCAGAGACCTGAAATTCTGCCATCCTGAACTCAAGAGTGGAGAATACTAG
1261    A  R  N  C  R  D  L  K  F  C  H  P  E  L  K  S  G  E  Y  *

             3850      3860      3870      3880      3890      3900
3841   GTTGACCCTAACCAAGGATGCAAATTGGATGCTATCAAGGTATTCTGTAATATGGAAACT
1281    V  D  P  N  Q  G  C  K  L  D  A  I  K  V  F  C  N  M  E  T

             3910      3920      3930      3940      3950      3960
3901   GGGGAAACATGCATAAGTGCCAATCCTTTGAATGTTCCACGGAAACACTGGTGGACAGAT
1301    G  E  T  C  I  S  A  N  P  L  N  V  P  R  K  H  W  W  T  D

             3970      3980      3990      4000      4010      4020
3961   TCTAGTGCTGAGAAGAAACACGTTTGGTTTGGAGAGTCCATGGATGGTGGTTTTCAGTTT
1321    S  S  A  E  K  K  H  V  W  F  G  E  S  M  D  G  G  F  Q  F

             4030      4040      4050      4060      4070      4080
4021   AGCTACGGCAATCCTGAACTTCCTGAAGATGTCCTTGATGTGCAGCTGGCATTCCTTCGA
1341    S  Y  G  N  P  E  L  P  E  D  V  L  D  V  Q  L  A  F  L  R

             4090      4100      4110      4120      4130      4140
4081   CTTCTCTCCAGCCGAGCTTCCCAGAACATCACATATCACTGCAAAAATAGCATTGCATAC
1361    L  L  S  S  R  A  S  Q  N  I  T  Y  H  C  K  N  S  I  A  Y

             4150      4160      4170      4180      4190      4200
4141   ATGGATCAGGCCAGTGGAAATGTAAAGAAGGCCCTGAAGCTGATGGGGTCAAATGAAGGT
1381    M  D  Q  A  S  G  N  V  K  K  A  L  K  L  M  G  S  N  E  G

             4210      4220      4230      4240      4250      4260
4201   GAATTCAAGGCTGAAGGAAATAGCAAATTCACCTACACAGTTCTGGAGGATGGTTGCACG
1401    E  F  K  A  E  G  N  S  K  F  T  Y  T  V  L  E  D  G  C  T

             4270      4280      4290      4300      4310      4320
4261   AAACACACTGGGGAATGGAGCAAAACAGTCTTTGAATATCGAACACGCAAGGCTGTGAGA
1421    K  H  T  G  E  W  S  K  T  V  F  E  Y  R  T  R  K  A  V  R

             4330      4340      4350      4360      4370      4380
4321   CTACCTATTGTAGATATTGCACCCTATGACATTGGTGGTCCTGATCAAGAATTTGGTGTG
1441    L  P  I  V  D  I  A  P  Y  D  I  G  G  P  D  Q  E  F  G  V

             4390      4400      4410      4420
4381   GACGTTGGCCCTGTTTGCTTTTTATAAGCGGCCGCTCGAG
1461    D  V  G  P  V  C  F  L  *        Xho-I
```

Gene & protein sequence of the **FANCC_W506X (pTS1162)** construct in the context of the pcDNA 3.1 vector, under control of the CMV promotor and the BGH polyA signal:

```
                10        20        30        40        50        60
1         GCTAGCATGGCTCAAGATTCAGTAGATCTTTCTTGTGATTATCAGTTTTGGATGCAGAAG
1         Nhe-I  M  A  Q  D  S  V  D  L  S  C  D  Y  Q  F  W  M  Q  K

               70        80        90       100       110       120
61        CTTTCTGTATGGGATCAGGCTTCCACTTTGGAAACCCAGCAAGACACCTGTCTTCACGTG
21         L  S  V  W  D  Q  A  S  T  L  E  T  Q  Q  D  T  C  L  H  V

              130       140       150       160       170       180
121       GCTCAGTTCCAGGAGTTCCTAAGGAAGATGTATGAAGCCTTGAAAGAGATGGATTCTAAT
41         A  Q  F  Q  E  F  L  R  K  M  Y  E  A  L  K  E  M  D  S  N

              190       200       210       220       230       240
181       ACAGTCATTGAAAGATTCCCCACAATTGGTCAACTGTTGGCAAAAGCTTGTTGGAATCCT
61         T  V  I  E  R  F  P  T  I  G  Q  L  L  A  K  A  C  W  N  P

              250       260       270       280       290       300
241       TTTATTTTAGCATATGATGAAAGCCAAAAAATTCTAATATGGTGCTTATGTTGTCTAATT
81         F  I  L  A  Y  D  E  S  Q  K  I  L  I  W  C  L  C  C  L  I

              310       320       330       340       350       360
301       AACAAAGAACCACAGAATTCTGGACAATCAAAACTTAACTCCTGGATACAGGGTGTATTA
101        N  K  E  P  Q  N  S  G  Q  S  K  L  N  S  W  I  Q  G  V  L

              370       380       390       400       410       420
361       TCTCATATACTTTCAGCACTCAGATTTGATAAAGAAGTTGCTCTTTTCACTCAAGGTCTT
121        S  H  I  L  S  A  L  R  F  D  K  E  V  A  L  F  T  Q  G  L

              430       440       450       460       470       480
421       GGGTATGCACCTATAGATTACTATCCTGGTTTGCTTAAAAATATGGTTTTATCATTAGCG
141        G  Y  A  P  I  D  Y  Y  P  G  L  L  K  N  M  V  L  S  L  A

              490       500       510       520       530       540
481       TCTGAACTCAGAGAGAATCATCTTAATGGATTTAACACTCAAAGGCGAATGGCTCCCGAG
161        S  E  L  R  E  N  H  L  N  G  F  N  T  Q  R  R  M  A  P  E

              550       560       570       580       590       600
541       CGAGTGGCGTCCCTGTCACGAGTTTGTGTCCCACTTATTACCCTGACAGATGTTGACCCC
181        R  V  A  S  L  S  R  V  C  V  P  L  I  T  L  T  D  V  D  P

              610       620       630       640       650       660
601       CTGGTGGAGGCTCTCCTCATCTGTCATGGACGTGAACCTCAGGAAATCCTCCAGCCAGAG
201        L  V  E  A  L  L  I  C  H  G  R  E  P  Q  E  I  L  Q  P  E

              670       680       690       700       710       720
661       TTCTTTGAGGCTGTAAACGAGGCCATTTTGCTGAAGAAGATTTCTCTCCCCATGTCAGCT
221        F  F  E  A  V  N  E  A  I  L  L  K  K  I  S  L  P  M  S  A

              730       740       750       760       770       780
721       GTAGTCTGCCTCTGGCTTCGGCACCTTCCCAGCCTTGAAAAAGCAATGCTGCATCTTTTT
241        V  V  C  L  W  L  R  H  L  P  S  L  E  K  A  M  L  H  L  F

              790       800       810       820       830       840
781       GAAAAGCTAATCTCCAGTGAGAGAAATTGTCTGAGAAGGATCGAATGCTTTATAAAAGAT
261        E  K  L  I  S  S  E  R  N  C  L  R  R  I  E  C  F  I  K  D

              850       860       870       880       890       900
841       TCATCGCTGCCTCAAGCAGCCTGCCACCCTGCCATATTCCGGGTTGTTGATGAGATGTTC
```

```
281        S  S  L  P  Q  A  A  C  H  P  A  I  F  R  V  V  D  E  M  F

                 910        920        930        940        950        960
901        AGGTGTGCACTCCTGGAAACCGATGGGGCCCTGGAAATCATAGCCACTATTCAGGTGTTT
301         R  C  A  L  L  E  T  D  G  A  L  E  I  I  A  T  I  Q  V  F

                 970        980        990       1000       1010       1020
961        ACGCAGTGCTTTGTAGAAGCTCTGGAGAAAGCAAGCAAGCAGCTGCGGTTTGCACTCAAG
321         T  Q  C  F  V  E  A  L  E  K  A  S  K  Q  L  R  F  A  L  K

               1030       1040       1050       1060       1070       1080
1021       ACCTACTTTCCTTACACTTCTCCATCTCTTGCCATGGTGCTGCTGCAAGACCCTCAAGAT
341         T  Y  F  P  Y  T  S  P  S  L  A  M  V  L  L  Q  D  P  Q  D

               1090       1100       1110       1120       1130       1140
1081       ATCCCTCGGGGACACTGGCTCCAGACACTGAAGCATATTTCTGAACTGCTCAGAGAAGCA
361         I  P  R  G  H  W  L  Q  T  L  K  H  I  S  E  L  L  R  E  A

               1150       1160       1170       1180       1190       1200
1141       GTTGAAGACCAGACTCATGGGTCCTGCGGAGGTCCCTTTGAGAGCTGGTTCCTGTTCATT
381         V  E  D  Q  T  H  G  S  C  G  G  P  F  E  S  W  F  L  F  I

               1210       1220       1230       1240       1250       1260
1201       CACTTCGGAGGATGGGCTGAGATGGTGGCAGAGCAATTACTGATGTCGGCAGCCGAACCC
401         H  F  G  G  W  A  E  M  V  A  E  Q  L  L  M  S  A  A  E  P

               1270       1280       1290       1300       1310       1320
1261       CCCACGGCCCTGCTGTGGCTCTTGGCCTTCTACTACGGCCCCCGTGATGGGAGGCAGCAG
421         P  T  A  L  L  W  L  L  A  F  Y  Y  G  P  R  D  G  R  Q  Q

               1330       1340       1350       1360       1370       1380
1321       AGAGCACAGACTATGGTCCAGGTGAAGGCCGTGCTGGGCCACCTCCTGGCAATGTCCAGA
441         R  A  Q  T  M  V  Q  V  K  A  V  L  G  H  L  L  A  M  S  R

               1390       1400       1410       1420       1430       1440
1381       AGCAGCAGCCTCTCAGCCCAGGACCTGCAGACGGTAGCAGGACAGGGCACAGACACAGAC
461         S  S  S  L  S  A  Q  D  L  Q  T  V  A  G  Q  G  T  D  T  D

               1450       1460       1470       1480       1490       1500
1441       CTCAGAGCTCCTGCACAACAGCTGATCAGGCACCTTCTCCTCAACTTCCTGCTCTGGGCT
481         L  R  A  P  A  Q  Q  L  I  R  H  L  L  L  N  F  L  L  W  A

               1510       1520       1530       1540       1550       1560
1501       CCTGGAGGCCACACGATCGCCTAGGATGTCATCACCCTGATGGCTCACACTGCTGAGATA
501         P  G  G  H  T  I  A  *  D  V  I  T  L  M  A  H  T  A  E  I

               1570       1580       1590       1600       1610       1620
1561       ACTCACGAGATCATTGGCTTTCTTGACCAGACCTTGTACAGATGGAATCGTCTTGGCATT
521         T  H  E  I  I  G  F  L  D  Q  T  L  Y  R  W  N  R  L  G  I

               1630       1640       1650       1660       1670       1680
1621       GAAAGCCCTAGATCAGAAAAACTGGCCCGAGAGCTCCTTAAAGAGCTGCGAACTCAAGTC
541         E  S  P  R  S  E  K  L  A  R  E  L  L  K  E  L  R  T  Q  V

               1690
1681       TAGGCGGCCGCTCGAG
561         *        Xho-I
```

Gene & protein sequence of the **IL2RG_W237X (pTS1165)** construct in the context of the pcDNA 3.1 vector, under control of the CMV promotor and the BGH polyA signal:

```
                10        20        30        40        50        60
1       GCTAGCATGTTGAAGCCATCATTACCATTCACATCCCTCTTATTCCTGCAGCTGCCCCTG
1       Nhe-I  M  L  K  P  S  L  P  F  T  S  L  L  F  L  Q  L  P  L

                70        80        90       100       110       120
61      CTGGGAGTGGGGCTGAACACGACAATTCTGACGCCCAATGGGAATGAAGACACCACAGCT
21       L  G  V  G  L  N  T  T  I  L  T  P  N  G  N  E  D  T  T  A

               130       140       150       160       170       180
121     GATTTCTTCCTGACCACTATGCCCACTGACTCCCTCAGTGTTTCCACTCTGCCCCTCCCA
41       D  F  F  L  T  T  M  P  T  D  S  L  S  V  S  T  L  P  L  P

               190       200       210       220       230       240
181     GAGGTTCAGTGTTTTGTGTTCAATGTCGAGTACATGAATTGCACTTGGAACAGCAGCTCT
61       E  V  Q  C  F  V  F  N  V  E  Y  M  N  C  T  W  N  S  S  S

               250       260       270       280       290       300
241     GAGCCCCAGCCTACCAACCTCACTCTGCATTATTGGTACAAGAACTCGGATAATGATAAA
81       E  P  Q  P  T  N  L  T  L  H  Y  W  Y  K  N  S  D  N  D  K

               310       320       330       340       350       360
301     GTCCAGAAGTGCAGCCACTATCTATTCTCTGAAGAAATCACTTCTGGCTGTCAGTTGCAA
101      V  Q  K  C  S  H  Y  L  F  S  E  E  I  T  S  G  C  Q  L  Q

               370       380       390       400       410       420
361     AAAAAGGAGATCCACCTCTACCAAACATTTGTTGTTCAGCTCCAGGACCCACGGGAACCC
121      K  K  E  I  H  L  Y  Q  T  F  V  V  Q  L  Q  D  P  R  E  P

               430       440       450       460       470       480
421     AGGAGACAGGCCACACAGATGCTAAAACTGCAGAATCTGGTGATCCCCTGGGCTCCAGAG
141      R  R  Q  A  T  Q  M  L  K  L  Q  N  L  V  I  P  W  A  P  E

               490       500       510       520       530       540
481     AACCTAACACTTCACAAACTGAGTGAATCCCAGCTAGAACTGAACTGGAACAACAGATTC
161      N  L  T  L  H  K  L  S  E  S  Q  L  E  L  N  W  N  N  R  F

               550       560       570       580       590       600
541     TTGAACCACTGTTTGGAGCACTTGGTGCAGTACCGGACTGACTGGGACCACAGCTGGACT
181      L  N  H  C  L  E  H  L  V  Q  Y  R  T  D  W  D  H  S  W  T

               610       620       630       640       650       660
601     GAACAATCAGTGGATTATAGACATAAGTTCTCCTTGCCTAGTGTGGATGGGCAGAAACGC
201      E  Q  S  V  D  Y  R  H  K  F  S  L  P  S  V  D  G  Q  K  R

               670       680       690       700       710       720
661     TACACGTTTCGTGTTCGGAGCCGCTTTAACCCACTCTGTGGAAGTGCTCAGCATTAGAGT
221      Y  T  F  R  V  R  S  R  F  N  P  L  C  G  S  A  Q  H  *  S

               730       740       750       760       770       780
721     GAATGGAGCCACCCAATCCACTGGGGGAGCAATACTTCAAAAGAGAATCCTTTCCTGTTT
241      E  W  S  H  P  I  H  W  G  S  N  T  S  K  E  N  P  F  L  F

               790       800       810       820       830       840
781     GCATTGGAAGCCGTGGTTATCTCTGTTGGCTCCATGGGATTGATTATCAGCCTTCTCTGT
261      A  L  E  A  V  V  I  S  V  G  S  M  G  L  I  I  S  L  L  C

               850       860       870       880       890       900
```

```
841       GTGTATTTCTGGCTGGAACGGACGATGCCCCGAATTCCCACCCTGAAGAACCTAGAGGAT
281        V  Y  F  W  L  E  R  T  M  P  R  I  P  T  L  K  N  L  E  D

                 910       920       930       940       950       960
901       CTTGTTACTGAATACCACGGGAACTTTTCGGCCTGGAGTGGTGTGTCTAAGGGACTGGCT
301        L  V  T  E  Y  H  G  N  F  S  A  W  S  G  V  S  K  G  L  A

                 970       980       990      1000      1010      1020
961       GAGAGTCTGCAGCCAGACTACAGTGAACGACTCTGCCTCGTCAGTGAGATTCCCCCAAAA
321        E  S  L  Q  P  D  Y  S  E  R  L  C  L  V  S  E  I  P  P  K

               1030      1040      1050      1060      1070      1080
1021      GGAGGGGCCCTTGGGGAGGGGCCTGGGGCCTCCCCATGCAACCAGCATAGCCCCTACTGG
341        G  G  A  L  G  E  G  P  G  A  S  P  C  N  Q  H  S  P  Y  W

               1090      1100      1110      1120
1081      GCCCCCCCATGTTACACCCTAAAGCCTGAAACCTGAGCGGCCGCTCGAG
361        A  P  P  C  Y  T  L  K  P  E  T  *        Xho-I
```

Gene & protein sequence of the **murine IDUA wt (pTS1113)** construct in the context of the pcDNA 3.1 vector, under control of the CMV promotor and the BGH polyA signal:

```
                10        20        30        40        50        60
1       GGATCCGCCACCATGCGACCCCCGCGTCCCTCCTCAGCTATGCTGACGTTTTTTGCTGCG
1       BamHI       M  R  P  P  R  P  S  S  A  M  L  T  F  F  A  A

                70        80        90       100       110       120
61      TTCTTGGCCGCGCCCTTGGCGCTGGCTGAGTCACCGTACCTGGTGCGTGTGGACGCAGCC
21       F  L  A  A  P  L  A  L  A  E  S  P  Y  L  V  R  V  D  A  A

               130       140       150       160       170       180
121     CGCCCGCTGAGGCCTCTGTTGCCCTTCTGGAGGAGCACCGGCTTCTGCCCCCCACTGCCT
41       R  P  L  R  P  L  L  P  F  W  R  S  T  G  F  C  P  P  L  P

               190       200       210       220       230       240
181     CACGACCAGGCTGACCAGTACGACCTTAGTTGGGACCAGCAACTGAACCTTGCCTACATA
61       H  D  Q  A  D  Q  Y  D  L  S  W  D  Q  Q  L  N  L  A  Y  I

               250       260       270       280       290       300
241     GGTGCCGTACCTCACAGTGGCATTGAGCAGGTCCGGATACACTGGCTGCTGGATCTCATC
81       G  A  V  P  H  S  G  I  E  Q  V  R  I  H  W  L  L  D  L  I

               310       320       330       340       350       360
301     ACAGCCAGGAAGTCACCTGGGCAGGGACTTATGTACAACTTCACCCACTTGGATGCATTC
101      T  A  R  K  S  P  G  Q  G  L  M  Y  N  F  T  H  L  D  A  F

               370       380       390       400       410       420
361     TTGGACCTTCTCATGGAGAACCAGCTTCTCCCTGGATTTGAGCTCATGGGCAGTCCTTCT
121      L  D  L  L  M  E  N  Q  L  L  P  G  F  E  L  M  G  S  P  S

               430       440       450       460       470       480
421     GGGTACTTCACGGACTTTGATGACAAGCAGCAGGTGTTTGAATGGAAGGACCTGGTTTCT
141      G  Y  F  T  D  F  D  D  K  Q  Q  V  F  E  W  K  D  L  V  S

               490       500       510       520       530       540
481     CTCTTGGCCAGGAGATACATTGGTAGGTATGGGCTGACACACGTTTCCAAGTGGAACTTT
161      L  L  A  R  R  Y  I  G  R  Y  G  L  T  H  V  S  K  W  N  F

               550       560       570       580       590       600
541     GAGACTTGGAATGAACCAGACCACCATGACTTTGACAACGTGTCCATGACCACACAAGGC
181      E  T  W  N  E  P  D  H  H  D  F  D  N  V  S  M  T  T  Q  G

               610       620       630       640       650       660
601     TTCCTGAATTACTATGATGCCTGCTCTGAGGGGCTGCGCATTGCCAGCCCCACTTTGAAG
201      F  L  N  Y  Y  D  A  C  S  E  G  L  R  I  A  S  P  T  L  K

               670       680       690       700       710       720
661     TTGGGTGGTCCTGGGGATTCCTTCCACCCCCTGCCAAGGTCACCAATGTGCTGGAGCCTC
221      L  G  G  P  G  D  S  F  H  P  L  P  R  S  P  M  C  W  S  L

               730       740       750       760       770       780
721     CTGGGTCACTGTGCCAATGGAACCAACTTCTTCACTGGCGAGGTGGGCGTGCGTCTGGAT
241      L  G  H  C  A  N  G  T  N  F  F  T  G  E  V  G  V  R  L  D

               790       800       810       820       830       840
781     TACATCTCCCTGCACAAGAAGGGTGCAGGTAGCTCCATCGCCATCCTGGAGCAGGAGATG
261      Y  I  S  L  H  K  K  G  A  G  S  S  I  A  I  L  E  Q  E  M

               850       860       870       880       890       900
```

```
841       GCAGTTGTGGAGCAGGTCCAGCAGCTCTTCCCTGAGTTCAAGGATACCCCTATTTACAAT
281        A  V  V  E  Q  V  Q  Q  L  F  P  E  F  K  D  T  P  I  Y  N

                 910       920       930       940       950       960
901       GACGAGGCAGACCCTCTGGTGGGCTGGTCCCTGCCACAACCTTGGAGAGCTGATGTGACT
301        D  E  A  D  P  L  V  G  W  S  L  P  Q  P  W  R  A  D  V  T

                 970       980       990      1000      1010      1020
961       TATGCGGCCCTGGTGGTGAAGGTCATTGCACAGCACCAGAACCTGCTGTTTGCCAACAGC
321        Y  A  A  L  V  V  K  V  I  A  Q  H  Q  N  L  L  F  A  N  S

               1030      1040      1050      1060      1070      1080
1021      AGTTCCTCCATGCGCTATGTGCTCCTCAGCAATGACAATGCCTTCCTGAGCTACCACCCG
341        S  S  S  M  R  Y  V  L  L  S  N  D  N  A  F  L  S  Y  H  P

               1090      1100      1110      1120      1130      1140
1081      TACCCTTTCTCCCAGCGCACACTTACTGCTCGATTCCAGGTCAACAATACTCACCCACCC
361        Y  P  F  S  Q  R  T  L  T  A  R  F  Q  V  N  N  T  H  P  P

               1150      1160      1170      1180      1190      1200
1141      CACGTGCAGTTGCTGCGAAAGCCAGTACTCACAGTCATGGGGCTCATGGCCCTGTTGGAT
381        H  V  Q  L  L  R  K  P  V  L  T  V  M  G  L  M  A  L  L  D

               1210      1220      1230      1240      1250      1260
1201      GGAGAACAACTCTGGGCAGAGGTCTCAAAGGCTGGGGCTGTGTTGGACAGCAATCATACA
401        G  E  Q  L  W  A  E  V  S  K  A  G  A  V  L  D  S  N  H  T

               1270      1280      1290      1300      1310      1320
1261      GTGGGTGTCCTGGCCAGCACCCATCACCCTGAAGGCTCCGCAGCGGCCTGGAGTACCACA
421        V  G  V  L  A  S  T  H  H  P  E  G  S  A  A  A  W  S  T  T

               1330      1340      1350      1360      1370      1380
1321      GTCCTCATCTACACTAGTGATGACACCCACGCACACCCCAACCACAGTATCCCTGTGACT
441        V  L  I  Y  T  S  D  D  T  H  A  H  P  N  H  S  I  P  V  T

               1390      1400      1410      1420      1430      1440
1381      CTTCGCCTGCGTGGGGTACCTCCTGGCTTGGATCTTGTCTACATAGTACTCTACTTAGAC
461        L  R  L  R  G  V  P  P  G  L  D  L  V  Y  I  V  L  Y  L  D

               1450      1460      1470      1480      1490      1500
1441      AATCAACTCAGCAGCCCCTACAGTGCGTGGCAGCACATGGGCCAGCCAGTCTTCCCCTCT
481        N  Q  L  S  S  P  Y  S  A  W  Q  H  M  G  Q  P  V  F  P  S

               1510      1520      1530      1540      1550      1560
1501      GCAGAGCAGTTCCGACGTATGCGCATGGTGGAGGACCCCGTGGCTGAGGCACCACGCCCC
501        A  E  Q  F  R  R  M  R  M  V  E  D  P  V  A  E  A  P  R  P

               1570      1580      1590      1600      1610      1620
1561      TTTCCTGCTAGGGGCCGCCTGACCCTACACCGGAAGCTTCCGGTGCCATCACTCCTGCTG
521        F  P  A  R  G  R  L  T  L  H  R  K  L  P  V  P  S  L  L  L

               1630      1640      1650      1660      1670      1680
1621      GTGCATGTATGCACACGCCCCTTGAAGCCACCTGGGCAGGTCAGCCGGCTCCGTGCACTG
541        V  H  V  C  T  R  P  L  K  P  P  G  Q  V  S  R  L  R  A  L

               1690      1700      1710      1720      1730      1740
1681      CCCCTGACACATGGACAGCTGATTTTGGTCTGGTCAGATGAGCGTGTGGGCTCCAAGTGC
561        P  L  T  H  G  Q  L  I  L  V  W  S  D  E  R  V  G  S  K  C

               1750      1760      1770      1780      1790      1800
1741      CTGTGGACATATGAGATCCAGTTTTCCCAGAAAGGTGAAGAGTATGCCCCAATCAACAGG
```

```
581          L  W  T  Y  E  I  Q  F  S  Q  K  G  E  E  Y  A  P  I  N  R

                  1810       1820       1830       1840       1850       1860
1801       AGGCCGTCTACTTTTAACCTCTTTGTGTTCAGCCCAGACACAGCTGTGGTCTCTGGCTCC
601          R  P  S  T  F  N  L  F  V  F  S  P  D  T  A  V  V  S  G  S

                  1870       1880       1890       1900       1910       1920
1861       TACCGAGTTCGAGCATTGGATTACTGGGCCCGGCCAGGCCCCTTCTCCGACCCTGTGACT
621          Y  R  V  R  A  L  D  Y  W  A  R  P  G  P  F  S  D  P  V  T

                  1930       1940       1950
1921       TACCTGGATGTCCCTGCCTCATGAGCGGCCGC
641          Y  L  D  V  P  A  S  *  Not-I
```

Gene & protein sequence of the **murine IDUA W392X (pTS1114)** construct in the context of the pcDNA 3.1 vector, under control of the CMV promotor and the BGH polyA signal:

```
                10        20        30        40        50        60
1       GGATCCGCCACCATGCGACCCCCGCGTCCCTCCTCAGCTATGCTGACGTTTTTTGCTGCG
1       BamHI       M  R  P  P  R  P  S  S  A  M  L  T  F  F  A  A

                70        80        90       100       110       120
61      TTCTTGGCCGCGCCCTTGGCGCTGGCTGAGTCACCGTACCTGGTGCGTGTGGACGCAGCC
21       F  L  A  A  P  L  A  L  A  E  S  P  Y  L  V  R  V  D  A  A

               130       140       150       160       170       180
121     CGCCCGCTGAGGCCTCTGTTGCCCTTCTGGAGGAGCACCGGCTTCTGCCCCCCACTGCCT
41       R  P  L  R  P  L  L  P  F  W  R  S  T  G  F  C  P  P  L  P

               190       200       210       220       230       240
181     CACGACCAGGCTGACCAGTACGACCTTAGTTGGGACCAGCAACTGAACCTTGCCTACATA
61       H  D  Q  A  D  Q  Y  D  L  S  W  D  Q  Q  L  N  L  A  Y  I

               250       260       270       280       290       300
241     GGTGCCGTACCTCACAGTGGCATTGAGCAGGTCCGGATACACTGGCTGCTGGATCTCATC
81       G  A  V  P  H  S  G  I  E  Q  V  R  I  H  W  L  L  D  L  I

               310       320       330       340       350       360
301     ACAGCCAGGAAGTCACCTGGGCAGGGACTTATGTACAACTTCACCCACTTGGATGCATTC
101      T  A  R  K  S  P  G  Q  G  L  M  Y  N  F  T  H  L  D  A  F

               370       380       390       400       410       420
361     TTGGACCTTCTCATGGAGAACCAGCTTCTCCCTGGATTTGAGCTCATGGGCAGTCCTTCT
121      L  D  L  L  M  E  N  Q  L  L  P  G  F  E  L  M  G  S  P  S

               430       440       450       460       470       480
421     GGGTACTTCACGGACTTTGATGACAAGCAGCAGGTGTTTGAATGGAAGGACCTGGTTTCT
141      G  Y  F  T  D  F  D  D  K  Q  Q  V  F  E  W  K  D  L  V  S

               490       500       510       520       530       540
481     CTCTTGGCCAGGAGATACATTGGTAGGTATGGGCTGACACACGTTTCCAAGTGGAACTTT
161      L  L  A  R  R  Y  I  G  R  Y  G  L  T  H  V  S  K  W  N  F

               550       560       570       580       590       600
541     GAGACTTGGAATGAACCAGACCACCATGACTTTGACAACGTGTCCATGACCACACAAGGC
181      E  T  W  N  E  P  D  H  H  D  F  D  N  V  S  M  T  T  Q  G

               610       620       630       640       650       660
601     TTCCTGAATTACTATGATGCCTGCTCTGAGGGGCTGCGCATTGCCAGCCCCACTTTGAAG
201      F  L  N  Y  Y  D  A  C  S  E  G  L  R  I  A  S  P  T  L  K

               670       680       690       700       710       720
661     TTGGGTGGTCCTGGGGATTCCTTCCACCCCCTGCCAAGGTCACCAATGTGCTGGAGCCTC
221      L  G  G  P  G  D  S  F  H  P  L  P  R  S  P  M  C  W  S  L

               730       740       750       760       770       780
721     CTGGGTCACTGTGCCAATGGAACCAACTTCTTCACTGGCGAGGTGGGCGTGCGTCTGGAT
241      L  G  H  C  A  N  G  T  N  F  F  T  G  E  V  G  V  R  L  D

               790       800       810       820       830       840
781     TACATCTCCCTGCACAAGAAGGGTGCAGGTAGCTCCATCGCCATCCTGGAGCAGGAGATG
261      Y  I  S  L  H  K  K  G  A  G  S  S  I  A  I  L  E  Q  E  M
```

```
                850       860       870       880       890       900
841       GCAGTTGTGGAGCAGGTCCAGCAGCTCTTCCCTGAGTTCAAGGATACCCCTATTTACAAT
281        A  V  V  E  Q  V  Q  Q  L  F  P  E  F  K  D  T  P  I  Y  N

                910       920       930       940       950       960
901       GACGAGGCAGACCCTCTGGTGGGCTGGTCCCTGCCACAACCTTGGAGAGCTGATGTGACT
301        D  E  A  D  P  L  V  G  W  S  L  P  Q  P  W  R  A  D  V  T

                970       980       990      1000      1010      1020
961       TATGCGGCCCTGGTGGTGAAGGTCATTGCACAGCACCAGAACCTGCTGTTTGCCAACAGC
321        Y  A  A  L  V  V  K  V  I  A  Q  H  Q  N  L  L  F  A  N  S

               1030      1040      1050      1060      1070      1080
1021      AGTTCCTCCATGCGCTATGTGCTCCTCAGCAATGACAATGCCTTCCTGAGCTACCACCCG
341        S  S  S  M  R  Y  V  L  L  S  N  D  N  A  F  L  S  Y  H  P

               1090      1100      1110      1120      1130      1140
1081      TACCCTTTCTCCCAGCGCACACTTACTGCTCGATTCCAGGTCAACAATACTCACCCACCC
361        Y  P  F  S  Q  R  T  L  T  A  R  F  Q  V  N  N  T  H  P  P

               1150      1160      1170      1180      1190      1200
1141      CACGTGCAGTTGCTGCGAAAGCCAGTACTCACAGTCATGGGGCTCATGGCCCTGTTGGAT
381        H  V  Q  L  L  R  K  P  V  L  T  V  M  G  L  M  A  L  L  D

               1210      1220      1230      1240      1250      1260
1201      GGAGAACAACTCTAGGCAGAGGTCTCAAAGGCTGGGGCTGTGTTGGACAGCAATCATACA
401        G  E  Q  L  *  A  E  V  S  K  A  G  A  V  L  D  S  N  H  T

               1270      1280      1290      1300      1310      1320
1261      GTGGGTGTCCTGGCCAGCACCCATCACCCTGAAGGCTCCGCAGCGGCCTGGAGTACCACA
421        V  G  V  L  A  S  T  H  H  P  E  G  S  A  A  A  W  S  T  T

               1330      1340      1350      1360      1370      1380
1321      GTCCTCATCTACACTAGTGATGACACCCACGCACACCCCAACCACAGTATCCCTGTGACT
441        V  L  I  Y  T  S  D  D  T  H  A  H  P  N  H  S  I  P  V  T

               1390      1400      1410      1420      1430      1440
1381      CTTCGCCTGCGTGGGGTACCTCCTGGCTTGGATCTTGTCTACATAGTACTCTACTTAGAC
461        L  R  L  R  G  V  P  P  G  L  D  L  V  Y  I  V  L  Y  L  D

               1450      1460      1470      1480      1490      1500
1441      AATCAACTCAGCAGCCCCTACAGTGCGTGGCAGCACATGGGCCAGCCAGTCTTCCCCTCT
481        N  Q  L  S  S  P  Y  S  A  W  Q  H  M  G  Q  P  V  F  P  S

               1510      1520      1530      1540      1550      1560
1501      GCAGAGCAGTTCCGACGTATGCGCATGGTGGAGGACCCCGTGGCTGAGGCACCACGCCCC
501        A  E  Q  F  R  R  M  R  M  V  E  D  P  V  A  E  A  P  R  P

               1570      1580      1590      1600      1610      1620
1561      TTTCCTGCTAGGGGCCGCCTGACCCTACACCGGAAGCTTCCGGTGCCATCACTCCTGCTG
521        F  P  A  R  G  R  L  T  L  H  R  K  L  P  V  P  S  L  L  L

               1630      1640      1650      1660      1670      1680
1621      GTGCATGTATGCACACGCCCCTTGAAGCCACCTGGGCAGGTCAGCCGGCTCCGTGCACTG
541        V  H  V  C  T  R  P  L  K  P  P  G  Q  V  S  R  L  R  A  L

               1690      1700      1710      1720      1730      1740
1681      CCCCTGACACATGGACAGCTGATTTTGGTCTGGTCAGATGAGCGTGTGGGCTCCAAGTGC
561        P  L  T  H  G  Q  L  I  L  V  W  S  D  E  R  V  G  S  K  C
```

```
                1750      1760      1770      1780      1790      1800
1741       CTGTGGACATATGAGATCCAGTTTTCCCAGAAAGGTGAAGAGTATGCCCCAATCAACAGG
581         L  W  T  Y  E  I  Q  F  S  Q  K  G  E  E  Y  A  P  I  N  R

                1810      1820      1830      1840      1850      1860
1801       AGGCCGTCTACTTTTAACCTCTTTGTGTTCAGCCCAGACACAGCTGTGGTCTCTGGCTCC
601         R  P  S  T  F  N  L  F  V  F  S  P  D  T  A  V  V  S  G  S

                1870      1880      1890      1900      1910      1920
1861       TACCGAGTTCGAGCATTGGATTACTGGGCCCGGCCAGGCCCCTTCTCCGACCCTGTGACT
621         Y  R  V  R  A  L  D  Y  W  A  R  P  G  P  F  S  D  P  V  T

                1930      1940      1950
1921       TACCTGGATGTCCCTGCCTCATGAGCGGCCGC
641         Y  L  D  V  P  A  S  *  Not-I
```

Gene & protein sequence of the **MYBPC3_W1098X (pTS1166)** construct in the context of the pcDNA 3.1 vector, under control of the CMV promotor and the BGH polyA signal:

```
                10        20        30        40        50        60
1       GCTAGCATGCCTGAGCCGGGGAAGAAGCCAGTCTCAGCTTTTAGCAAGAAGCCACGGTCA
1       Nhe-I  M  P  E  P  G  K  K  P  V  S  A  F  S  K  K  P  R  S

                70        80        90       100       110       120
61      GTGGAAGTGGCCGCAGGCAGCCCTGCCGTGTTCGAGGCCGAGACAGAGCGGGCAGGAGTG
21       V  E  V  A  A  G  S  P  A  V  F  E  A  E  T  E  R  A  G  V

               130       140       150       160       170       180
121     AAGGTGCGCTGGCAGCGCGGAGGCAGTGACATCAGCGCCAGCAACAAGTACGGCCTGGCC
41       K  V  R  W  Q  R  G  G  S  D  I  S  A  S  N  K  Y  G  L  A

               190       200       210       220       230       240
181     ACAGAGGGCACACGGCATACGCTGACAGTGCGGGAAGTGGGCCCTGCCGACCAGGGATCT
61       T  E  G  T  R  H  T  L  T  V  R  E  V  G  P  A  D  Q  G  S

               250       260       270       280       290       300
241     TACGCAGTCATTGCTGGCTCCTCCAAGGTCAAGTTCGACCTCAAGGTCATAGAGGCAGAG
81       Y  A  V  I  A  G  S  S  K  V  K  F  D  L  K  V  I  E  A  E

               310       320       330       340       350       360
301     AAGGCAGAGCCCATGCTGGCCCCTGCCCCTGCCCCTGCTGAGGCCACTGGAGCCCCTGGA
101      K  A  E  P  M  L  A  P  A  P  A  P  A  E  A  T  G  A  P  G

               370       380       390       400       410       420
361     GAAGCCCCGGCCCCAGCCGCTGAGCTGGGAGAAAGTGCCCCAAGTCCCAAAGGGTCAAGC
121      E  A  P  A  P  A  A  E  L  G  E  S  A  P  S  P  K  G  S  S

               430       440       450       460       470       480
421     TCAGCAGCTCTCAATGGTCCTACCCCTGGAGCCCCCGATGACCCCATTGGCCTCTTCGTG
141      S  A  A  L  N  G  P  T  P  G  A  P  D  D  P  I  G  L  F  V

               490       500       510       520       530       540
481     ATGCGGCCACAGGATGGCGAGGTGACCGTGGGTGGCAGCATCACCTTCTCAGCCCGCGTG
161      M  R  P  Q  D  G  E  V  T  V  G  G  S  I  T  F  S  A  R  V

               550       560       570       580       590       600
541     GCCGGCGCCAGCCTCCTGAAGCCGCCTGTGGTCAAGTGGTTCAAGGGCAAATGGGTGGAC
181      A  G  A  S  L  L  K  P  P  V  V  K  W  F  K  G  K  W  V  D

               610       620       630       640       650       660
601     CTGAGCAGCAAGGTGGGCCAGCACCTGCAGCTGCACGACAGCTACGACCGCGCCAGCAAG
201      L  S  S  K  V  G  Q  H  L  Q  L  H  D  S  Y  D  R  A  S  K

               670       680       690       700       710       720
661     GTCTATCTGTTCGAGCTGCACATCACCGATGCCCAGCCTGCCTTCACTGGCAGCTACCGC
221      V  Y  L  F  E  L  H  I  T  D  A  Q  P  A  F  T  G  S  Y  R

               730       740       750       760       770       780
721     TGTGAGGTGTCCACCAAGGACAAATTTGACTGCTCCAACTTCAATCTCACTGTCCACGAG
241      C  E  V  S  T  K  D  K  F  D  C  S  N  F  N  L  T  V  H  E

               790       800       810       820       830       840
781     GCCATGGGCACCGGAGACCTGGACCTCCTATCAGCCTTCCGCCGCACGAGCCTGGCTGGA
261      A  M  G  T  G  D  L  D  L  L  S  A  F  R  R  T  S  L  A  G
```

```
                850       860       870       880       890       900
841       GGTGGTCGGCGGATCAGTGATAGCCATGAGGACACTGGGATTCTGGACTTCAGCTCACTG
281        G  G  R  R  I  S  D  S  H  E  D  T  G  I  L  D  F  S  S  L

                910       920       930       940       950       960
901       CTGAAAAAGAGAGACAGTTTCCGGACCCCGAGGGACTCGAAGCTGGAGGCACCAGCAGAG
301        L  K  K  R  D  S  F  R  T  P  R  D  S  K  L  E  A  P  A  E

                970       980       990      1000      1010      1020
961       GAGGACGTGTGGGAGATCCTACGGCAGGCACCCCCATCTGAGTACGAGCGCATCGCCTTC
321        E  D  V  W  E  I  L  R  Q  A  P  P  S  E  Y  E  R  I  A  F

               1030      1040      1050      1060      1070      1080
1021      CAGTACGGCGTCACTGACCTGCGCGGCATGCTAAAGAGGCTCAAGGGCATGAGGCGCGAT
341        Q  Y  G  V  T  D  L  R  G  M  L  K  R  L  K  G  M  R  R  D

               1090      1100      1110      1120      1130      1140
1081      GAGAAGAAGAGCACAGCCTTTCAGAAGAAGCTGGAGCCGGCCTACCAGGTGAGCAAAGGC
361        E  K  K  S  T  A  F  Q  K  K  L  E  P  A  Y  Q  V  S  K  G

               1150      1160      1170      1180      1190      1200
1141      CACAAGATCCGGCTGACCGTGGAACTGGCTGACCATGACGCTGAGGTCAAATGGCTCAAG
381        H  K  I  R  L  T  V  E  L  A  D  H  D  A  E  V  K  W  L  K

               1210      1220      1230      1240      1250      1260
1201      AATGGCCAGGAGATCCAGATGAGCGGCAGCAAGTACATCTTTGAGTCCATCGGTGCCAAG
401        N  G  Q  E  I  Q  M  S  G  S  K  Y  I  F  E  S  I  G  A  K

               1270      1280      1290      1300      1310      1320
1261      CGTACCCTGACCATCAGCCAGTGCTCATTGGCGGACGACGCAGCCTACCAGTGCGTGGTG
421        R  T  L  T  I  S  Q  C  S  L  A  D  D  A  A  Y  Q  C  V  V

               1330      1340      1350      1360      1370      1380
1321      GGTGGCGAGAAGTGTAGCACGGAGCTCTTTGTGAAAGAGCCCCCTGTGCTCATCACGCGC
441        G  G  E  K  C  S  T  E  L  F  V  K  E  P  P  V  L  I  T  R

               1390      1400      1410      1420      1430      1440
1381      CCCTTGGAGGACCAGCTGGTGATGGTGGGGCAGCGGGTGGAGTTTGAGTGTGAAGTATCG
461        P  L  E  D  Q  L  V  M  V  G  Q  R  V  E  F  E  C  E  V  S

               1450      1460      1470      1480      1490      1500
1441      GAGGAGGGGGCGCAAGTCAAATGGCTGAAGGACGGGGTGGAGCTGACCCGGGAGGAGACC
481        E  E  G  A  Q  V  K  W  L  K  D  G  V  E  L  T  R  E  E  T

               1510      1520      1530      1540      1550      1560
1501      TTCAAATACCGGTTCAAGAAGGACGGGCAGAGACACCACCTGATCATCAACGAGGCCATG
501        F  K  Y  R  F  K  K  D  G  Q  R  H  H  L  I  I  N  E  A  M

               1570      1580      1590      1600      1610      1620
1561      CTGGAGGACGCGGGGCACTATGCACTGTGCACTAGCGGGGGCCAGGCGCTGGCTGAGCTC
521        L  E  D  A  G  H  Y  A  L  C  T  S  G  G  Q  A  L  A  E  L

               1630      1640      1650      1660      1670      1680
1621      ATTGTGCAGGAAAAGAAGCTGGAGGTGTACCAGAGCATCGCAGACCTGATGGTGGGCGCA
541        I  V  Q  E  K  K  L  E  V  Y  Q  S  I  A  D  L  M  V  G  A

               1690      1700      1710      1720      1730      1740
1681      AAGGACCAGGCGGTGTTCAAATGTGAGGTCTCAGATGAGAATGTTCGGGGGTGTGTGGCTG
561        K  D  Q  A  V  F  K  C  E  V  S  D  E  N  V  R  G  V  W  L

               1750      1760      1770      1780      1790      1800
```

```
1741      AAGAATGGGAAGGAGCTGGTGCCCGACAGCCGCATAAAGGTGTCCCACATCGGGCGGGTC
581        K  N  G  K  E  L  V  P  D  S  R  I  K  V  S  H  I  G  R  V

               1810      1820      1830      1840      1850      1860
1801      CACAAACTGACCATTGACGACGTCACACCTGCCGACGAGGCTGACTACAGCTTTGTGCCC
601        H  K  L  T  I  D  D  V  T  P  A  D  E  A  D  Y  S  F  V  P

               1870      1880      1890      1900      1910      1920
1861      GAGGGCTTCGCCTGCAACCTGTCAGCCAAGCTCCACTTCATGGAGGTCAAGATTGACTTC
621        E  G  F  A  C  N  L  S  A  K  L  H  F  M  E  V  K  I  D  F

               1930      1940      1950      1960      1970      1980
1921      GTACCCAGGCAGGAACCTCCCAAGATCCACCTGGACTGCCCAGGCCGCATACCAGACACC
641        V  P  R  Q  E  P  P  K  I  H  L  D  C  P  G  R  I  P  D  T

               1990      2000      2010      2020      2030      2040
1981      ATTGTGGTTGTAGCTGGAAATAAGCTACGTCTGGACGTCCCTATCTCTGGGGACCCTGCT
661        I  V  V  V  A  G  N  K  L  R  L  D  V  P  I  S  G  D  P  A

               2050      2060      2070      2080      2090      2100
2041      CCCACTGTGATCTGGCAGAAGGCTATCACGCAGGGGAATAAGGCCCCAGCCAGGCCAGCC
681        P  T  V  I  W  Q  K  A  I  T  Q  G  N  K  A  P  A  R  P  A

               2110      2120      2130      2140      2150      2160
2101      CCAGATGCCCCAGAGGACACAGGTGACAGCGATGAGTGGGTGTTTGACAAGAAGCTGCTG
701        P  D  A  P  E  D  T  G  D  S  D  E  W  V  F  D  K  K  L  L

               2170      2180      2190      2200      2210      2220
2161      TGTGAGACCGAGGGCCGGGTCCGCGTGGAGACCACCAAGGACCGCAGCATCTTCACGGTC
721        C  E  T  E  G  R  V  R  V  E  T  T  K  D  R  S  I  F  T  V

               2230      2240      2250      2260      2270      2280
2221      GAGGGGGCAGAGAAGGAAGATGAGGGCGTCTACACGGTCACAGTGAAGAACCCTGTGGGC
741        E  G  A  E  K  E  D  E  G  V  Y  T  V  T  V  K  N  P  V  G

               2290      2300      2310      2320      2330      2340
2281      GAGGACCAGGTCAACCTCACAGTCAAGGTCATCGACGTGCCAGACGCACCTGCGGCCCCC
761        E  D  Q  V  N  L  T  V  K  V  I  D  V  P  D  A  P  A  A  P

               2350      2360      2370      2380      2390      2400
2341      AAGATCAGCAACGTGGGAGAGGACTCCTGCACAGTACAGTGGGAGCCGCCTGCCTACGAT
781        K  I  S  N  V  G  E  D  S  C  T  V  Q  W  E  P  P  A  Y  D

               2410      2420      2430      2440      2450      2460
2401      GGCGGGCAGCCCATCCTGGGCTACATCCTGGAGCGCAAGAAGAAGAAGAGCTACCGGTGG
801        G  G  Q  P  I  L  G  Y  I  L  E  R  K  K  K  K  S  Y  R  W

               2470      2480      2490      2500      2510      2520
2461      ATGCGGCTGAACTTCGACCTGATTCAGGAGCTGAGTCATGAAGCGCGGCGCATGATCGAG
821        M  R  L  N  F  D  L  I  Q  E  L  S  H  E  A  R  R  M  I  E

               2530      2540      2550      2560      2570      2580
2521      GGCGTGGTGTACGAGATGCGCGTCTACGCGGTCAACGCCATCGGCATGTCCAGGCCCAGC
841        G  V  V  Y  E  M  R  V  Y  A  V  N  A  I  G  M  S  R  P  S

               2590      2600      2610      2620      2630      2640
2581      CCTGCCTCCCAGCCCTTCATGCCTATCGGTCCCCCCAGCGAACCCACCCACCTGGCAGTA
861        P  A  S  Q  P  F  M  P  I  G  P  P  S  E  P  T  H  L  A  V

               2650      2660      2670      2680      2690      2700
2641      GAGGACGTCTCTGACACCACGGTCTCCCTCAAGTGGCGGCCCCCAGAGCGCGTGGGAGCA
```

```
881        E  D  V  S  D  T  T  V  S  L  K  W  R  P  P  E  R  V  G  A

                2710       2720       2730       2740       2750       2760
2701       GGAGGCCTGGATGGCTACAGCGTGGAGTACTGCCCAGAGGGCTGCTCAGAGTGGGTGGCT
901         G  G  L  D  G  Y  S  V  E  Y  C  P  E  G  C  S  E  W  V  A

                2770       2780       2790       2800       2810       2820
2761       GCCCTGCAGGGGCTGACAGAGCACACATCGATACTGGTGAAGGACCTGCCCACGGGGGCC
921         A  L  Q  G  L  T  E  H  T  S  I  L  V  K  D  L  P  T  G  A

                2830       2840       2850       2860       2870       2880
2821       CGGCTGCTTTTCCGAGTGCGGGCACACAATATGGCAGGGCCTGGAGCCCCTGTTACCACC
941         R  L  L  F  R  V  R  A  H  N  M  A  G  P  G  A  P  V  T  T

                2890       2900       2910       2920       2930       2940
2881       ACGGAGCCGGTGACAGTGCAGGAGATCCTGCAACGGCCACGGCTTCAGCTGCCCAGGCAC
961         T  E  P  V  T  V  Q  E  I  L  Q  R  P  R  L  Q  L  P  R  H

                2950       2960       2970       2980       2990       3000
2941       CTGCGCCAGACCATTCAGAAGAAGGTCGGGGAGCCTGTGAACCTTCTCATCCCTTTCCAG
981         L  R  Q  T  I  Q  K  K  V  G  E  P  V  N  L  L  I  P  F  Q

                3010       3020       3030       3040       3050       3060
3001       GGCAAGCCCCGGCCTCAGGTGACCTGGACCAAAGAGGGGCAGCCCCTGGCAGGCGAGGAG
1001        G  K  P  R  P  Q  V  T  W  T  K  E  G  Q  P  L  A  G  E  E

                3070       3080       3090       3100       3110       3120
3061       GTGAGCATCCGCAACAGCCCCACAGACACCATCCTGTTCATCCGGGCCGCTCGCCGCGTG
1021        V  S  I  R  N  S  P  T  D  T  I  L  F  I  R  A  A  R  R  V

                3130       3140       3150       3160       3170       3180
3121       CATTCAGGCACTTACCAGGTGACGGTGCGCATTGAGAACATGGAGGACAAGGCCACGCTG
1041        H  S  G  T  Y  Q  V  T  V  R  I  E  N  M  E  D  K  A  T  L

                3190       3200       3210       3220       3230       3240
3181       GTGCTGCAGGTTGTTGACAAGCCAAGTCCTCCCCAGGATCTCCGGGTGACTGACGCCTGG
1061        V  L  Q  V  V  D  K  P  S  P  P  Q  D  L  R  V  T  D  A  W

                3250       3260       3270       3280       3290       3300
3241       GGTCTTAATGTGGCTCTGGAGTGGAAGCCACCCCAGGATGTCGGCAACACGGAGCTCTAG
1081        G  L  N  V  A  L  E  W  K  P  P  Q  D  V  G  N  T  E  L  *

                3310       3320       3330       3340       3350       3360
3301       GGGTACACAGTGCAGAAAGCCGACAAGAAGACCATGGAGTGGTTCACCGTCTTGGAGCAT
1101        G  Y  T  V  Q  K  A  D  K  K  T  M  E  W  F  T  V  L  E  H

                3370       3380       3390       3400       3410       3420
3361       TACCGCCGCACCCACTGCGTGGTGCCAGAGCTCATCATTGGCAATGGCTACTACTTCCGC
1121        Y  R  R  T  H  C  V  V  P  E  L  I  I  G  N  G  Y  Y  F  R

                3430       3440       3450       3460       3470       3480
3421       GTCTTCAGCCAGAATATGGTTGGCTTTAGTGACAGAGCGGCCACCACCAAGGAGCCCGTC
1141        V  F  S  Q  N  M  V  G  F  S  D  R  A  A  T  T  K  E  P  V

                3490       3500       3510       3520       3530       3540
3481       TTTATCCCCAGACCAGGCATCACCTATGAGCCACCCAACTATAAGGCCCTGGACTTCTCC
1161        F  I  P  R  P  G  I  T  Y  E  P  P  N  Y  K  A  L  D  F  S

                3550       3560       3570       3580       3590       3600
3541       GAGGCCCCAAGCTTCACCCAGCCCCTGGTGAACCGCTCGGTCATCGCGGGCTACACTGCT
1181        E  A  P  S  F  T  Q  P  L  V  N  R  S  V  I  A  G  Y  T  A
```

```
              3610      3620      3630      3640      3650      3660
3601      ATGCTCTGCTGTGCTGTCCGGGGTAGCCCCAAGCCCAAGATTTCCTGGTTCAAGAATGGC
1201       M  L  C  C  A  V  R  G  S  P  K  P  K  I  S  W  F  K  N  G

              3670      3680      3690      3700      3710      3720
3661      CTGGACCTGGGAGAAGACGCCCGCTTCCGCATGTTCAGCAAGCAGGGAGTGTTGACTCTG
1221       L  D  L  G  E  D  A  R  F  R  M  F  S  K  Q  G  V  L  T  L

              3730      3740      3750      3760      3770      3780
3721      GAGATTAGAAAGCCCTGCCCCTTTGACGGGGGCATCTATGTCTGCAGGGCCACCAACTTA
1241       E  I  R  K  P  C  P  F  D  G  G  I  Y  V  C  R  A  T  N  L

              3790      3800      3810      3820      3830      3840
3781      CAGGGCGAGGCACGGTGTGAGTGCCGCCTGGAGGTGCGAGTGCCTCAGTGAGCGGCCGCT
1261       Q  G  E  A  R  C  E  C  R  L  E  V  R  V  P  Q  *

3841      CGAG
1281      Xho-I
```

Gene & protein sequence of the **PINK1_W437X (pTS65)** construct in the context of the pcDNA6 v5-His vector, under control of the CMV promotor and the BGH polyA signal:

```
                10        20        30        40        50        60
1       GCTAGCATGGCGGTGCGACAGGCGCTGGGCCGCGGCCTGCAGCTGGGTCGAGCGCTGCTG
1       Nhe-I  M  A  V  R  Q  A  L  G  R  G  L  Q  L  G  R  A  L  L

                70        80        90       100       110       120
61      CTGCGCTTCACGGGCAAGCCCGGCCGGGCCTACGGCTTGGGGCGGCCGGGCCCGGCGGCG
21       L  R  F  T  G  K  P  G  R  A  Y  G  L  G  R  P  G  P  A  A

               130       140       150       160       170       180
121     GGCTGTGTCCGCGGGGAGCGTCCAGGCTGGGCCGCAGGACCGGGCGCGGAGCCTCGCAGG
41       G  C  V  R  G  E  R  P  G  W  A  A  G  P  G  A  E  P  R  R

               190       200       210       220       230       240
181     GTCGGGCTCGGGCTCCCTAACCGTCTCCGCTTCTTCCGCCAGTCGGTGGCCGGGCTGGCG
61       V  G  L  G  L  P  N  R  L  R  F  F  R  Q  S  V  A  G  L  A

               250       260       270       280       290       300
241     GCGCGGTTGCAGCGGCAGTTCGTGGTGCGGGCCTGGGGCTGCGCGGGCCCTTGCGGCCGG
81       A  R  L  Q  R  Q  F  V  V  R  A  W  G  C  A  G  P  C  G  R

               310       320       330       340       350       360
301     GCAGTCTTTCTGGCCTTCGGGCTAGGGCTGGGCCTCATCGAGGAAAAACAGGCGGAGAGC
101      A  V  F  L  A  F  G  L  G  L  G  L  I  E  E  K  Q  A  E  S

               370       380       390       400       410       420
361     CGGCGGGCGGTCTCGGCCTGTCAGGAGATCCAGGCAATTTTTACCCAGAAAAGCAAGCCG
121      R  R  A  V  S  A  C  Q  E  I  Q  A  I  F  T  Q  K  S  K  P

               430       440       450       460       470       480
421     GGGCCTGACCCGTTGGACACGAGACGCTTGCAGGGCTTTCGGCTGGAGGAGTATCTGATA
141      G  P  D  P  L  D  T  R  R  L  Q  G  F  R  L  E  E  Y  L  I

               490       500       510       520       530       540
481     GGGCAGTCCATTGGTAAGGGCTGCAGTGCTGCTGTGTATGAAGCCACCATGCCTACATTG
161      G  Q  S  I  G  K  G  C  S  A  A  V  Y  E  A  T  M  P  T  L

               550       560       570       580       590       600
541     CCCCAGAACCTGGAGGTGACAAAGAGCACCGGGTTGCTTCCAGGGAGAGGCCCAGGTACC
181      P  Q  N  L  E  V  T  K  S  T  G  L  L  P  G  R  G  P  G  T

               610       620       630       640       650       660
601     AGTGCACCAGGAGAAGGGCAGGAGCGAGCTCCGGGGGCCCCTGCCTTCCCCTTGGCCATC
201      S  A  P  G  E  G  Q  E  R  A  P  G  A  P  A  F  P  L  A  I

               670       680       690       700       710       720
661     AAGATGATGTGGAACATCTCGGCAGGTTCCTCCAGCGAAGCCATCTTGAACACAATGAGC
221      K  M  M  W  N  I  S  A  G  S  S  S  E  A  I  L  N  T  M  S

               730       740       750       760       770       780
721     CAGGAGCTGGTCCCAGCGAGCCGAGTGGCCTTGGCTGGGGAGTATGGAGCAGTCACTTAC
241      Q  E  L  V  P  A  S  R  V  A  L  A  G  E  Y  G  A  V  T  Y

               790       800       810       820       830       840
781     AGAAAATCCAAGAGAGGTCCCAAGCAACTAGCCCCTCACCCCAACATCATCCGGGTTCTC
261      R  K  S  K  R  G  P  K  Q  L  A  P  H  P  N  I  I  R  V  L
```

```
                 850       860       870       880       890       900
841       CGCGCCTTCACCTCTTCCGTGCCGCTGCTGCCAGGGGCCCTGGTCGACTACCCTGATGTG
281        R  A  F  T  S  S  V  P  L  L  P  G  A  L  V  D  Y  P  D  V

                 910       920       930       940       950       960
901       CTGCCCTCACGCCTCCACCCTGAAGGCCTGGGCCATGGCCGGACGCTCTTTCTAGTCATG
301        L  P  S  R  L  H  P  E  G  L  G  H  G  R  T  L  F  L  V  M

                 970       980       990      1000      1010      1020
961       AAGAACTATCCCTGTACCCTGCGCCAGTACCTTTGTGTGAACACACCCAGCCCCCGCCTC
321        K  N  Y  P  C  T  L  R  Q  Y  L  C  V  N  T  P  S  P  R  L

               1030      1040      1050      1060      1070      1080
1021      GCCGCCATGATGCTGCTGCAGCTGCTGGAAGGCGTGGACCATCTGGTTCAACAGGGCATC
341        A  A  M  M  L  L  Q  L  L  E  G  V  D  H  L  V  Q  Q  G  I

               1090      1100      1110      1120      1130      1140
1081      GCGCACAGAGACCTGAAATCCGACAACATCCTTGTGGAGCTGGACCCAGACGGCTGCCCC
361        A  H  R  D  L  K  S  D  N  I  L  V  E  L  D  P  D  G  C  P

               1150      1160      1170      1180      1190      1200
1141      TGGCTGGTGATCGCAGATTTTGGCTGCTGCCTGGCTGATGAGAGCATCGGCCTGCAGTTG
381        W  L  V  I  A  D  F  G  C  C  L  A  D  E  S  I  G  L  Q  L

               1210      1220      1230      1240      1250      1260
1201      CCCTTCAGCAGCTGGTACGTGGATCGGGGCGGAAACGGCTGTCTGATGGCCCCAGAGGTG
401        P  F  S  S  W  Y  V  D  R  G  G  N  G  C  L  M  A  P  E  V

               1270      1280      1290      1300      1310      1320
1261      TCCACGGCCCGTCCTGGCCCCAGGGCAGTGATTGACTACAGCAAGGCTGATGCCTAGGCA
421        S  T  A  R  P  G  P  R  A  V  I  D  Y  S  K  A  D  A  *  A

               1330      1340      1350      1360      1370      1380
1321      GTGGGAGCCATCGCCTATGAAATCTTCGGGCTTGTCAATCCCTTCTACGGCCAGGGCAAG
441        V  G  A  I  A  Y  E  I  F  G  L  V  N  P  F  Y  G  Q  G  K

               1390      1400      1410      1420      1430      1440
1381      GCCCACCTTGAAAGCCGCAGCTACCAAGAGGCTCAGCTACCTGCACTGCCCGAGTCAGTG
461        A  H  L  E  S  R  S  Y  Q  E  A  Q  L  P  A  L  P  E  S  V

               1450      1460      1470      1480      1490      1500
1441      CCTCCAGACGTGAGACAGTTGGTGAGGGCACTGCTCCAGCGAGAGGCCAGCAAGAGACCA
481        P  P  D  V  R  Q  L  V  R  A  L  L  Q  R  E  A  S  K  R  P

               1510      1520      1530      1540      1550      1560
1501      TCTGCCCGAGTAGCCGCAAATGTGCTTCATCTAAGCCTCTGGGGTGAACATATTCTAGCC
501        S  A  R  V  A  A  N  V  L  H  L  S  L  W  G  E  H  I  L  A

               1570      1580      1590      1600      1610      1620
1561      CTGAAGAATCTGAAGTTAGACAAGATGGTTGGCTGGCTCCTCCAACAATCGGCCGCCACT
521        L  K  N  L  K  L  D  K  M  V  G  W  L  L  Q  Q  S  A  A  T

               1630      1640      1650      1660      1670      1680
1621      TTGTTGGCCAACAGGCTCACAGAGAAGTGTTGTGTGGAAACAAAAATGAAGATGCTCTTT
541        L  L  A  N  R  L  T  E  K  C  C  V  E  T  K  M  K  M  L  F

               1690      1700      1710      1720      1730      1740
1681      CTGGCTAACCTGGAGTGTGAAACGCTCTGCCAGGCAGCCCTCCTCCTCTGCTCATGGAGG
561        L  A  N  L  E  C  E  T  L  C  Q  A  A  L  L  L  C  S  W  R

               1750      1760
```

1741 GCAGCCCTGCTCGAG**TCTAGA**
581 A A L L E **Xho-I**

Gene & protein sequence of the **dual-luciferase wt/wt** (Firefly luciferase wild-type, 2A peptide, Renilla luciferase wild-type) construct in the context of the pShuttle-CMV (pTS554) or the pcDNA 3.1 (pTS656) vector, under control of the CMV promotor and either the BGH- (pTS554) or the SV40- (pTS656) polyA signal:

```
                10        20        30        40        50        60
1       AAGCTTCTAGATAAGATATGACTTCGAAAGTTTATGATCCAGAACAAAGGAAACGGATGA
1       Hind-III           M  T  S  K  V  Y  D  P  E  Q  R  K  R  M

                70        80        90       100       110       120
61      TAACTGGTCCGCAGTGGTGGGCCAGATGTAAACAAATGAATGTTCTTGATTCATTTATTA
20      I  T  G  P  Q  W  W  A  R  C  K  Q  M  N  V  L  D  S  F  I

               130       140       150       160       170       180
121     ATTATTATGATTCAGAAAAACATGCAGAAAATGCTGTTATTTTTTTACATGGTAACGCGG
40      N  Y  Y  D  S  E  K  H  A  E  N  A  V  I  F  L  H  G  N  A

               190       200       210       220       230       240
181     CCTCTTCTTATTTATGGCGACATGTTGTGCCACATATTGAGCCAGTAGCGCGGTGTATTA
60      A  S  S  Y  L  W  R  H  V  V  P  H  I  E  P  V  A  R  C  I

               250       260       270       280       290       300
241     TACCAGACCTTATTGGTATGGGCAAATCAGGCAAATCTGGTAATGGTTCTTATAGGTTAC
80      I  P  D  L  I  G  M  G  K  S  G  K  S  G  N  G  S  Y  R  L

               310       320       330       340       350       360
301     TTGATCATTACAAATATCTTACTGCATGGTTTGAACTTCTTAATTTACCAAAGAAGATCA
100     L  D  H  Y  K  Y  L  T  A  W  F  E  L  L  N  L  P  K  K  I

               370       380       390       400       410       420
361     TTTTTGTCGGCCATGATTGGGGTGCTTGTTTGGCATTTCATTATAGCTATGAGCATCAAG
120     I  F  V  G  H  D  W  G  A  C  L  A  F  H  Y  S  Y  E  H  Q

               430       440       450       460       470       480
421     ATAAGATCAAAGCAATAGTTCACGCTGAAAGTGTAGTAGATGTGATTGAATCATGGGATG
140     D  K  I  K  A  I  V  H  A  E  S  V  V  D  V  I  E  S  W  D

               490       500       510       520       530       540
481     AATGGCCTGATATTGAAGAAGATATTGCGTTGATCAAATCTGAAGAAGGAGAAAAAATGG
160     E  W  P  D  I  E  E  D  I  A  L  I  K  S  E  E  G  E  K  M

               550       560       570       580       590       600
541     TTTTGGAGAATAACTTCTTCGTGGAAACCATGTTGCCATCAAAAATCATGAGAAAGTTAG
180     V  L  E  N  N  F  F  V  E  T  M  L  P  S  K  I  M  R  K  L

               610       620       630       640       650       660
601     AACCAGAAGAATTTGCAGCATATCTTGAACCATTCAAAGAGAAAGGTGAAGTTCGTCGTC
200     E  P  E  E  F  A  A  Y  L  E  P  F  K  E  K  G  E  V  R  R

               670       680       690       700       710       720
661     CAACATTATCATGGCCTCGTGAAATCCCGTTAGTAAAAGGTGGTAAACCTGACGTTGTAC
220     P  T  L  S  W  P  R  E  I  P  L  V  K  G  G  K  P  D  V  V

               730       740       750       760       770       780
721     AAATTGTTAGGAATTATAATGCTTATCTACGTGCAAGTGATGATTTACCAAAAATGTTTA
240     Q  I  V  R  N  Y  N  A  Y  L  R  A  S  D  D  L  P  K  M  F

               790       800       810       820       830       840
781     TTGAATCGGACCCAGGATTCTTTTCCAATGCTATTGTTGAAGGTGCCAAGAAGTTTCCTA
260     I  E  S  D  P  G  F  F  S  N  A  I  V  E  G  A  K  K  F  P
```

```
                850       860       870       880       890       900
841       ATACTGAATTTGTCAAAGTAAAAGGTCTTCATTTTTCGCAAGAAGATGCACCTGATGAAA
280       N  T  E  F  V  K  V  K  G  L  H  F  S  Q  E  D  A  P  D  E

                910       920       930       940       950       960
901       TGGGAAAATATATCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAACAAGGAAGCGGAG
300       M  G  K  Y  I  K  S  F  V  E  R  V  L  K  N  E  Q  G  S  G

                970       980       990      1000      1010      1020
961       CTACTAACTTCAGCCTGCTGAAGCAGGCTGGAGACGTGGAGGAGAACCCTGGACCTATGG
320       A  T  N  F  S  L  L  K  Q  A  G  D  V  E  E  N  P  G  P  M

               1030      1040      1050      1060      1070      1080
1021      AAGATGCCAAAAACATTAAGAAGGGCCCAGCGCCATTCTACCCACTCGAAGACGGGACCG
340       E  D  A  K  N  I  K  K  G  P  A  P  F  Y  P  L  E  D  G  T

               1090      1100      1110      1120      1130      1140
1081      CCGGCGAGCAGCTGCACAAAGCCATGAAGCGCTACGCCCTGGTGCCCGGCACCATCGCCT
360       A  G  E  Q  L  H  K  A  M  K  R  Y  A  L  V  P  G  T  I  A

               1150      1160      1170      1180      1190      1200
1141      TTACCGACGCACATATCGAGGTGGACATTACCTACGCCGAGTACTTCGAGATGAGCGTTC
380       F  T  D  A  H  I  E  V  D  I  T  Y  A  E  Y  F  E  M  S  V

               1210      1220      1230      1240      1250      1260
1201      GGCTGGCAGAAGCTATGAAGCGCTATGGGCTGAATACAAACCATCGGATCGTGGTGTGCA
400       R  L  A  E  A  M  K  R  Y  G  L  N  T  N  H  R  I  V  V  C

               1270      1280      1290      1300      1310      1320
1261      GCGAGAATAGCTTGCAGTTCTTCATGCCCGTGTTGGGTGCCCTGTTCATCGGTGTGGCTG
420       S  E  N  S  L  Q  F  F  M  P  V  L  G  A  L  F  I  G  V  A

               1330      1340      1350      1360      1370      1380
1321      TGGCCCCAGCTAACGACATCTACAACGAGCGCGAGCTGCTGAACAGCATGGGCATCAGCC
440       V  A  P  A  N  D  I  Y  N  E  R  E  L  L  N  S  M  G  I  S

               1390      1400      1410      1420      1430      1440
1381      AGCCCACCGTCGTATTCGTGAGCAAGAAAGGGCTGCAAAAGATCCTCAACGTGCAAAAGA
460       Q  P  T  V  V  F  V  S  K  K  G  L  Q  K  I  L  N  V  Q  K

               1450      1460      1470      1480      1490      1500
1441      AGCTACCGATCATACAAAAGATCATCATCATGGATAGCAAGACCGACTACCAGGGCTTCC
480       K  L  P  I  I  Q  K  I  I  I  M  D  S  K  T  D  Y  Q  G  F

               1510      1520      1530      1540      1550      1560
1501      AAAGCATGTACACCTTCGTGACTTCCCATTTGCCACCCGGCTTCAACGAGTACGACTTCG
500       Q  S  M  Y  T  F  V  T  S  H  L  P  P  G  F  N  E  Y  D  F

               1570      1580      1590      1600      1610      1620
1561      TGCCCGAGAGCTTCGACCGGGACAAAACCATCGCCCTGATCATGAACAGTAGTGGCAGTA
520       V  P  E  S  F  D  R  D  K  T  I  A  L  I  M  N  S  S  G  S

               1630      1640      1650      1660      1670      1680
1621      CCGGATTGCCCAAGGGCGTAGCCCTACCGCACCGCACCGCTTGTGTCCGATTCAGTCATG
540       T  G  L  P  K  G  V  A  L  P  H  R  T  A  C  V  R  F  S  H

               1690      1700      1710      1720      1730      1740
1681      CCCGCGACCCCATCTTCGGCAACCAGATCATCCCCGACACCGCTATCCTCAGCGTGGTGC
560       A  R  D  P  I  F  G  N  Q  I  I  P  D  T  A  I  L  S  V  V
```

```
                1750      1760      1770      1780      1790      1800
1741      CATTTCACCACGGCTTCGGCATGTTCACCACGCTGGGCTACTTGATCTGCGGCTTTCGGG
580       P  F  H  H  G  F  G  M  F  T  T  L  G  Y  L  I  C  G  F  R

                1810      1820      1830      1840      1850      1860
1801      TCGTGCTCATGTACCGCTTCGAGGAGGAGCTATTCTTGCGCAGCTTGCAAGACTATAAGA
600       V  V  L  M  Y  R  F  E  E  E  L  F  L  R  S  L  Q  D  Y  K

                1870      1880      1890      1900      1910      1920
1861      TTCAATCTGCCCTGCTGGTGCCCACACTATTTAGCTTCTTCGCTAAGAGCACTCTCATCG
620       I  Q  S  A  L  L  V  P  T  L  F  S  F  F  A  K  S  T  L  I

                1930      1940      1950      1960      1970      1980
1921      ACAAGTACGACCTAAGCAACTTGCACGAGATCGCCAGCGGCGGGGCGCCGCTCAGCAAGG
640       D  K  Y  D  L  S  N  L  H  E  I  A  S  G  G  A  P  L  S  K

                1990      2000      2010      2020      2030      2040
1981      AGGTAGGTGAGGCCGTGGCCAAACGCTTCCACCTACCAGGCATCCGCCAGGGCTACGGCC
660       E  V  G  E  A  V  A  K  R  F  H  L  P  G  I  R  Q  G  Y  G

                2050      2060      2070      2080      2090      2100
2041      TGACAGAAACAACCAGCGCCATTCTGATCACCCCCGAAGGGGACGACAAGCCTGGCGCAG
680       L  T  E  T  T  S  A  I  L  I  T  P  E  G  D  D  K  P  G  A

                2110      2120      2130      2140      2150      2160
2101      TAGGCAAGGTGGTGCCCTTCTTCGAGGCTAAGGTGGTGGACTTGGACACCGGTAAGACAC
700       V  G  K  V  V  P  F  F  E  A  K  V  V  D  L  D  T  G  K  T

                2170      2180      2190      2200      2210      2220
2161      TGGGTGTGAACCAGCGCGGCGAGCTGTGCGTCCGTGGCCCCATGATCATGAGCGGCTACG
720       L  G  V  N  Q  R  G  E  L  C  V  R  G  P  M  I  M  S  G  Y

                2230      2240      2250      2260      2270      2280
2221      TTAACAACCCCGAGGCTACAAACGCTCTCATCGACAAGGACGGCTGGCTGCACAGCGGCG
740       V  N  N  P  E  A  T  N  A  L  I  D  K  D  G  W  L  H  S  G

                2290      2300      2310      2320      2330      2340
2281      ACATCGCCTACTGGGACGAGGACGAGCACTTCTTCATCGTGGACCGGCTGAAGAGCCTGA
760       D  I  A  Y  W  D  E  D  E  H  F  F  I  V  D  R  L  K  S  L

                2350      2360      2370      2380      2390      2400
2341      TCAAATACAAGGGCTACCAGGTAGCCCCAGCCGAACTGGAGAGCATCCTGCTGCAACACC
780       I  K  Y  K  G  Y  Q  V  A  P  A  E  L  E  S  I  L  L  Q  H

                2410      2420      2430      2440      2450      2460
2401      CCAACATCTTCGACGCCGGGGTCGCCGGCCTGCCCGACGACGATGCCGGCGAGCTGCCCG
800       P  N  I  F  D  A  G  V  A  G  L  P  D  D  D  A  G  E  L  P

                2470      2480      2490      2500      2510      2520
2461      CCGCAGTCGTCGTGCTGGAACACGGTAAAACCATGACCGAGAAGGAGATCGTGGACTATG
820       A  A  V  V  V  L  E  H  G  K  T  M  T  E  K  E  I  V  D  Y

                2530      2540      2550      2560      2570      2580
2521      TGGCCAGCCAGGTCACAACCGCCAAGAAGCTGCGCGGTGGTGTTGTGTTCGTGGACGAGG
840       V  A  S  Q  V  T  T  A  K  K  L  R  G  G  V  V  F  V  D  E

                2590      2600      2610      2620      2630      2640
2581      TGCCTAAAGGACTGACCGGCAAGTTGGACGCCCGCAAGATCCGCGAGATTCTCATTAAGG
860       V  P  K  G  L  T  G  K  L  D  A  R  K  I  R  E  I  L  I  K

                2650      2660      2670
```

```
2641      CCAAGAAGGGCGGCAAGATCGCCGTGTAAGAATTC
880       A  K  K  G  G  K  I  A  V  *  EcoRI
```

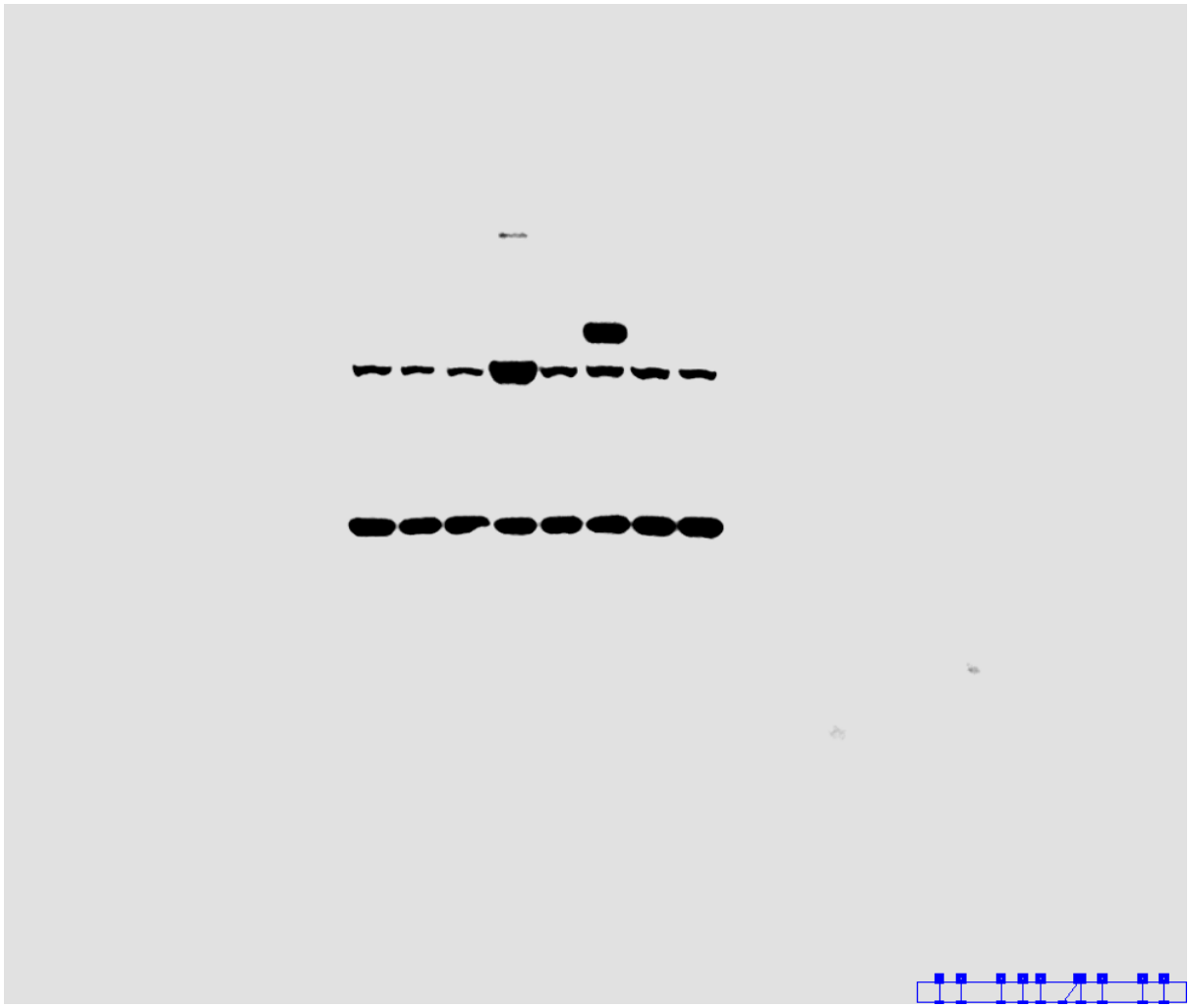
Gene & protein sequence of the **dual-luciferase wt/amb** (Firefly luciferase wild-type, 2A peptide, Renilla luciferase W417X amber) construct in the context of the pShuttle (pTS555) or the pcDNA 3.1 (pTS657) vector, under control of the CMV promotor and either the BGH- (pTS555) or the SV40- (pTS657) polyA signal:

```
                10        20        30        40        50        60
1       AAGCTTCTAGATAAGATATGACTTCGAAAGTTTATGATCCAGAACAAAGGAAACGGATGA
1       Hind-III           M  T  S  K  V  Y  D  P  E  Q  R  K  R  M

                70        80        90       100       110       120
61      TAACTGGTCCGCAGTGGTGGGCCAGATGTAAACAAATGAATGTTCTTGATTCATTTATTA
20      I  T  G  P  Q  W  W  A  R  C  K  Q  M  N  V  L  D  S  F  I

               130       140       150       160       170       180
121     ATTATTATGATTCAGAAAAACATGCAGAAAATGCTGTTATTTTTTTACATGGTAACGCGG
40      N  Y  Y  D  S  E  K  H  A  E  N  A  V  I  F  L  H  G  N  A

               190       200       210       220       230       240
181     CCTCTTCTTATTTATGGCGACATGTTGTGCCACATATTGAGCCAGTAGCGCGGTGTATTA
60      A  S  S  Y  L  W  R  H  V  V  P  H  I  E  P  V  A  R  C  I

               250       260       270       280       290       300
241     TACCAGACCTTATTGGTATGGGCAAATCAGGCAAATCTGGTAATGGTTCTTATAGGTTAC
80      I  P  D  L  I  G  M  G  K  S  G  K  S  G  N  G  S  Y  R  L

               310       320       330       340       350       360
301     TTGATCATTACAAATATCTTACTGCATGGTTTGAACTTCTTAATTTACCAAAGAAGATCA
100     L  D  H  Y  K  Y  L  T  A  W  F  E  L  L  N  L  P  K  K  I

               370       380       390       400       410       420
361     TTTTTGTCGGCCATGATTGGGGTGCTTGTTTGGCATTTCATTATAGCTATGAGCATCAAG
120     I  F  V  G  H  D  W  G  A  C  L  A  F  H  Y  S  Y  E  H  Q

               430       440       450       460       470       480
421     ATAAGATCAAAGCAATAGTTCACGCTGAAAGTGTAGTAGATGTGATTGAATCATGGGATG
140     D  K  I  K  A  I  V  H  A  E  S  V  V  D  V  I  E  S  W  D

               490       500       510       520       530       540
481     AATGGCCTGATATTGAAGAAGATATTGCGTTGATCAAATCTGAAGAAGGAGAAAAAATGG
160     E  W  P  D  I  E  E  D  I  A  L  I  K  S  E  E  G  E  K  M

               550       560       570       580       590       600
541     TTTTGGAGAATAACTTCTTCGTGGAAACCATGTTGCCATCAAAAATCATGAGAAAGTTAG
180     V  L  E  N  N  F  F  V  E  T  M  L  P  S  K  I  M  R  K  L

               610       620       630       640       650       660
601     AACCAGAAGAATTTGCAGCATATCTTGAACCATTCAAAGAGAAAGGTGAAGTTCGTCGTC
200     E  P  E  E  F  A  A  Y  L  E  P  F  K  E  K  G  E  V  R  R

               670       680       690       700       710       720
661     CAACATTATCATGGCCTCGTGAAATCCCGTTAGTAAAAGGTGGTAAACCTGACGTTGTAC
220     P  T  L  S  W  P  R  E  I  P  L  V  K  G  G  K  P  D  V  V

               730       740       750       760       770       780
721     AAATTGTTAGGAATTATAATGCTTATCTACGTGCAAGTGATGATTTACCAAAAATGTTTA
240     Q  I  V  R  N  Y  N  A  Y  L  R  A  S  D  D  L  P  K  M  F

               790       800       810       820       830       840
781     TTGAATCGGACCCAGGATTCTTTTCCAATGCTATTGTTGAAGGTGCCAAGAAGTTTCCTA
260     I  E  S  D  P  G  F  F  S  N  A  I  V  E  G  A  K  K  F  P
```

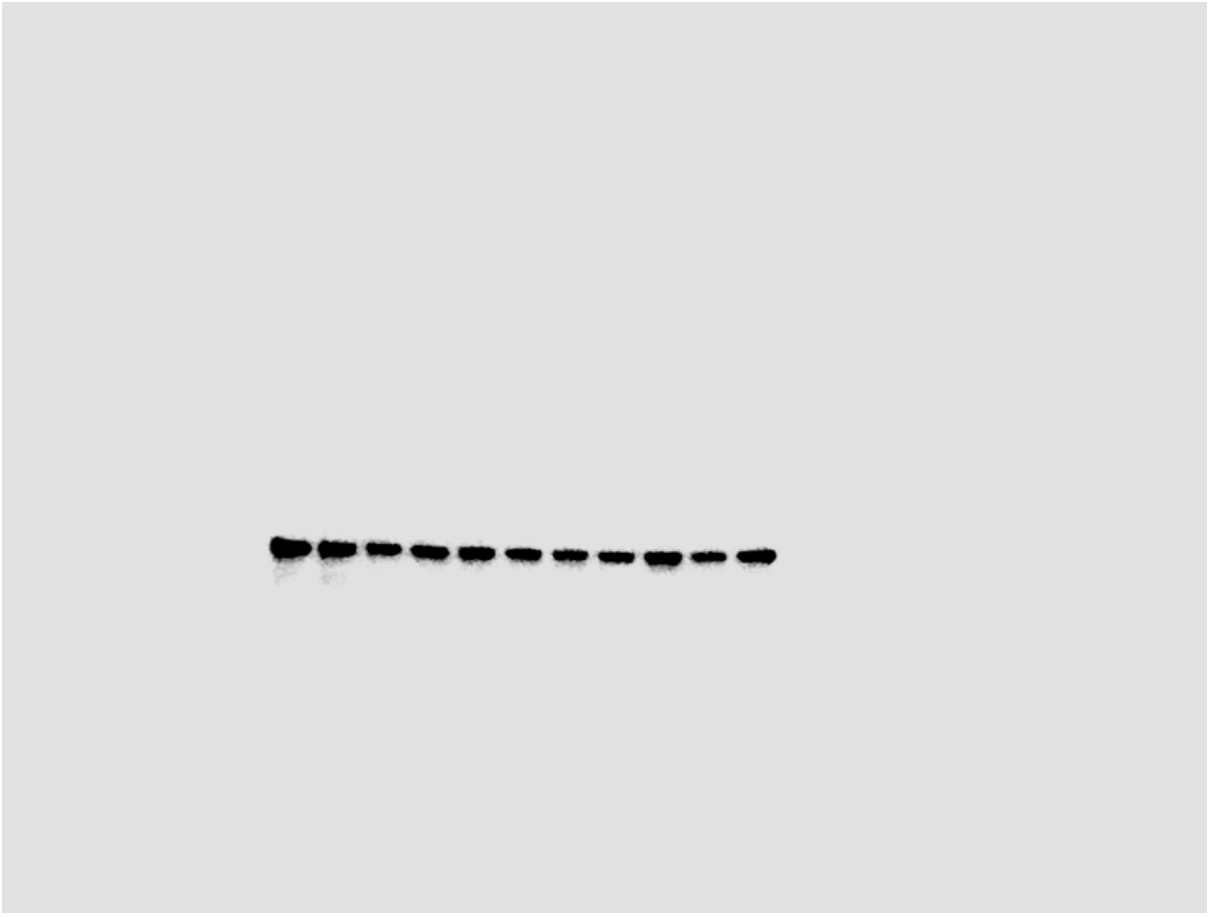
```
                850       860       870       880       890       900
841       ATACTGAATTTGTCAAAGTAAAAGGTCTTCATTTTTCGCAAGAAGATGCACCTGATGAAA
280       N  T  E  F  V  K  V  K  G  L  H  F  S  Q  E  D  A  P  D  E

                910       920       930       940       950       960
901       TGGGAAAATATATCAAATCGTTCGTTGAGCGAGTTCTCAAAAATGAACAAGGAAGCGGAG
300       M  G  K  Y  I  K  S  F  V  E  R  V  L  K  N  E  Q  G  S  G

                970       980       990      1000      1010      1020
961       CTACTAACTTCAGCCTGCTGAAGCAGGCTGGAGACGTGGAGGAGAACCCTGGACCTATGG
320       A  T  N  F  S  L  L  K  Q  A  G  D  V  E  E  N  P  G  P  M

               1030      1040      1050      1060      1070      1080
1021      AAGATGCCAAAAACATTAAGAAGGGCCCAGCGCCATTCTACCCACTCGAAGACGGGACCG
340       E  D  A  K  N  I  K  K  G  P  A  P  F  Y  P  L  E  D  G  T

               1090      1100      1110      1120      1130      1140
1081      CCGGCGAGCAGCTGCACAAAGCCATGAAGCGCTACGCCCTGGTGCCCGGCACCATCGCCT
360       A  G  E  Q  L  H  K  A  M  K  R  Y  A  L  V  P  G  T  I  A

               1150      1160      1170      1180      1190      1200
1141      TTACCGACGCACATATCGAGGTGGACATTACCTACGCCGAGTACTTCGAGATGAGCGTTC
380       F  T  D  A  H  I  E  V  D  I  T  Y  A  E  Y  F  E  M  S  V

               1210      1220      1230      1240      1250      1260
1201      GGCTGGCAGAAGCTATGAAGCGCTATGGGCTGAATACAAACCATCGGATCGTGGTGTGCA
400       R  L  A  E  A  M  K  R  Y  G  L  N  T  N  H  R  I  V  V  C

               1270      1280      1290      1300      1310      1320
1261      GCGAGAATAGCTTGCAGTTCTTCATGCCCGTGTTGGGTGCCCTGTTCATCGGTGTGGCTG
420       S  E  N  S  L  Q  F  F  M  P  V  L  G  A  L  F  I  G  V  A

               1330      1340      1350      1360      1370      1380
1321      TGGCCCCAGCTAACGACATCTACAACGAGCGCGAGCTGCTGAACAGCATGGGCATCAGCC
440       V  A  P  A  N  D  I  Y  N  E  R  E  L  L  N  S  M  G  I  S

               1390      1400      1410      1420      1430      1440
1381      AGCCCACCGTCGTATTCGTGAGCAAGAAAGGGCTGCAAAAGATCCTCAACGTGCAAAAGA
460       Q  P  T  V  V  F  V  S  K  K  G  L  Q  K  I  L  N  V  Q  K

               1450      1460      1470      1480      1490      1500
1441      AGCTACCGATCATACAAAAGATCATCATCATGGATAGCAAGACCGACTACCAGGGCTTCC
480       K  L  P  I  I  Q  K  I  I  I  M  D  S  K  T  D  Y  Q  G  F

               1510      1520      1530      1540      1550      1560
1501      AAAGCATGTACACCTTCGTGACTTCCCATTTGCCACCCGGCTTCAACGAGTACGACTTCG
500       Q  S  M  Y  T  F  V  T  S  H  L  P  P  G  F  N  E  Y  D  F

               1570      1580      1590      1600      1610      1620
1561      TGCCCGAGAGCTTCGACCGGGACAAAACCATCGCCCTGATCATGAACAGTAGTGGCAGTA
520       V  P  E  S  F  D  R  D  K  T  I  A  L  I  M  N  S  S  G  S

               1630      1640      1650      1660      1670      1680
1621      CCGGATTGCCCAAGGGCGTAGCCCTACCGCACCGCACCGCTTGTGTCCGATTCAGTCATG
540       T  G  L  P  K  G  V  A  L  P  H  R  T  A  C  V  R  F  S  H

               1690      1700      1710      1720      1730      1740
1681      CCCGCGACCCCATCTTCGGCAACCAGATCATCCCCGACACCGCTATCCTCAGCGTGGTGC
560       A  R  D  P  I  F  G  N  Q  I  I  P  D  T  A  I  L  S  V  V
```

```
                1750      1760      1770      1780      1790      1800
1741      CATTTCACCACGGCTTCGGCATGTTCACCACGCTGGGCTACTTGATCTGCGGCTTTCGGG
580       P  F  H  H  G  F  G  M  F  T  T  L  G  Y  L  I  C  G  F  R

                1810      1820      1830      1840      1850      1860
1801      TCGTGCTCATGTACCGCTTCGAGGAGGAGCTATTCTTGCGCAGCTTGCAAGACTATAAGA
600       V  V  L  M  Y  R  F  E  E  E  L  F  L  R  S  L  Q  D  Y  K

                1870      1880      1890      1900      1910      1920
1861      TTCAATCTGCCCTGCTGGTGCCCACACTATTTAGCTTCTTCGCTAAGAGCACTCTCATCG
620       I  Q  S  A  L  L  V  P  T  L  F  S  F  F  A  K  S  T  L  I

                1930      1940      1950      1960      1970      1980
1921      ACAAGTACGACCTAAGCAACTTGCACGAGATCGCCAGCGGCGGGGCGCCGCTCAGCAAGG
640       D  K  Y  D  L  S  N  L  H  E  I  A  S  G  G  A  P  L  S  K

                1990      2000      2010      2020      2030      2040
1981      AGGTAGGTGAGGCCGTGGCCAAACGCTTCCACCTACCAGGCATCCGCCAGGGCTACGGCC
660       E  V  G  E  A  V  A  K  R  F  H  L  P  G  I  R  Q  G  Y  G

                2050      2060      2070      2080      2090      2100
2041      TGACAGAAACAACCAGCGCCATTCTGATCACCCCCGAAGGGGACGACAAGCCTGGCGCAG
680       L  T  E  T  T  S  A  I  L  I  T  P  E  G  D  D  K  P  G  A

                2110      2120      2130      2140      2150      2160
2101      TAGGCAAGGTGGTGCCCTTCTTCGAGGCTAAGGTGGTGGACTTGGACACCGGTAAGACAC
700       V  G  K  V  V  P  F  F  E  A  K  V  V  D  L  D  T  G  K  T

                2170      2180      2190      2200      2210      2220
2161      TGGGTGTGAACCAGCGCGGCGAGCTGTGCGTCCGTGGCCCCATGATCATGAGCGGCTACG
720       L  G  V  N  Q  R  G  E  L  C  V  R  G  P  M  I  M  S  G  Y

                2230      2240      2250      2260      2270      2280
2221      TTAACAACCCCGAGGCTACAAACGCTCTCATCGACAAGGACGGCTAGCTGCACAGCGGCG
740       V  N  N  P  E  A  T  N  A  L  I  D  K  D  G  *  L  H  S  G

                2290      2300      2310      2320      2330      2340
2281      ACATCGCCTACTGGGACGAGGACGAGCACTTCTTCATCGTGGACCGGCTGAAGAGCCTGA
760       D  I  A  Y  W  D  E  D  E  H  F  F  I  V  D  R  L  K  S  L

                2350      2360      2370      2380      2390      2400
2341      TCAAATACAAGGGCTACCAGGTAGCCCCAGCCGAACTGGAGAGCATCCTGCTGCAACACC
780       I  K  Y  K  G  Y  Q  V  A  P  A  E  L  E  S  I  L  L  Q  H

                2410      2420      2430      2440      2450      2460
2401      CCAACATCTTCGACGCCGGGGTCGCCGGCCTGCCCGACGACGATGCCGGCGAGCTGCCCG
800       P  N  I  F  D  A  G  V  A  G  L  P  D  D  D  A  G  E  L  P

                2470      2480      2490      2500      2510      2520
2461      CCGCAGTCGTCGTGCTGGAACACGGTAAAACCATGACCGAGAAGGAGATCGTGGACTATG
820       A  A  V  V  V  L  E  H  G  K  T  M  T  E  K  E  I  V  D  Y

                2530      2540      2550      2560      2570      2580
2521      TGGCCAGCCAGGTCACAACCGCCAAGAAGCTGCGCGGTGGTGTTGTGTTCGTGGACGAGG
840       V  A  S  Q  V  T  T  A  K  K  L  R  G  G  V  V  F  V  D  E

                2590      2600      2610      2620      2630      2640
2581      TGCCTAAAGGACTGACCGGCAAGTTGGACGCCCGCAAGATCCGCGAGATTCTCATTAAGG
860       V  P  K  G  L  T  G  K  L  D  A  R  K  I  R  E  I  L  I  K

                2650      2660      2670
```

```
2641      CCAAGAAGGGCGGCAAGATCGCCGTGTAAGAATTC
880       A  K  K  G  G  K  I  A  V  *  EcoRI
```



Uncropped blot used for Figure S11.



Uncropped blot used for ADAR1 part of Figure S19 (scan of the upper part of the cut membrane).

Uncropped blot used for β-Actin part of Figure S19 (scan of the lower part of the cut membrane).