Aus der Klink für Gastroenterologie, Hepatologie und Infektiologie

der Heinrich-Heine-Universität Düsseldorf

Klinischer Direktor: Univ.-Prof. Dr. Tom Lüdde

HIV-restriction factor APOBEC3C variants fused to Cas9 show single base editing activity

Dissertation

zur Erlangung des Grades eines Doktors der Medizin

der Medizinischen Fakultät der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Catherina Chiapella

2025

Als Inauguraldissertation gedruckt mit der Genehmigung der

Medizinischen Fakultät der Heinrich-Heine-Universität Düsseldorf

gez.:

Dekan: Prof. Dr. med. Nikolaj Klöcker

Erstgutachter: Prof. Dr. rer. nat. Carsten Münk

Zweitgutachter: Prof. Dr. med. Ingo Drexler

Dedicated to my grandparents

Zusammenfassung

Humane APOBEC3 (A3) repräsentieren eine Proteinfamilie aus sieben Cytidin-Desaminasen, A3A-D, F-H. Es handelt sich hierbei um Restriktionsfaktoren des angeborenen Immunsystems, die den Wirt vor einer Infektion mit Retroviren und DNA-Viren schützen sollen: Durch die Desaminierung von Desoxy-Cytidin (dC)-Resten zu Desoxy-Uridin (dU)-Resten des viralen Genoms, werden Hypermutationen eingefügt, die zu einer reduzierten Integration der retroviralen DNA in das Wirtsgenom führen und somit die Virusreplikation hemmen.

Aufgrund ihrer katalytischen Aktivität als Desaminase, werden A3-Proteine auch im Zusammenhang mit *Base Editing (BE),* einer Weiterentwicklung der CRISPR-Cas9-Methode, angewandt. Im Gegensatz zur herkömmlichen CRISPR-Cas9-Methode, die auf DNA-Doppelstrangbrüchen basiert, fügen *BE* Komplexe Punktmutationen durch Desaminasen ein. Uracil-Glykosylase-Inhibitoren (UGI) und enzymatisch deaktiviertes Cas9 (nCas9) ermöglichen den Erhalt der Punktmutation. Zur Erkennung des Editierungsziels innerhalb der DNA wird eine *single guide RNA* (*sgRNA*) benötigt.

Innerhalb der A3-Proteinfamilie werden besonders A3A, A3B und A3G für *BE* genutzt, humanes A3C (huA3C) zeigte hingegen bis jetzt nur geringe Effizienz. Im Rahmen dieser Arbeit wurden sechs verschiedene *BE* Komplexe mittels eines fluoreszenz-basierten Echtzeit-Reporter-Assays getestet. Der A3A-Komplex diente als positive Kontrolle, bei den restlichen fünf *BE* Konstrukten handelte es sich um huA3C und davon abgeleitete Varianten. Ziel war es einen effizienten A3C-*BE* Komplex zu identifizieren und charakterisieren. Hierfür wurden zunächst HEK293T Zellen mit den *BE* Komplexen, einer *sgRNA* und dem Reporter Plasmid transfiziert. *BE* Aktivität konnte sowohl mikroskopisch als auch mittels Durchflusszytometrie beobachtet werden. A3C Chimär 2 (A3C CH2) zeigte hohe Effizienz im Rahmen dieser Experimente, die restlichen vier A3C Desaminasen hingegen nur geringe bis keine *BE* Aktivität. Bei A3C CH2 handelt es sich um einen Chimär aus huA3C mit Sooty mangabey A3C (smmA3C)-like. Die ersten 37 Aminosäuren stimmen mit smmA3C-like überein, die restlichen Aminosäuren stammen von huA3C.

Es wurden verschiedene Theorien aufgestellt, um die hohe *BE* Effizienz von A3C CH2 zu erklären. Fehlerhafte Proteinfaltung der anderen Varianten, Unterschiede der intrazellulären Lokalisierung oder eine stärkere Affinität zu nCas9 des A3C CH2 wurden diskutiert. Es konnten im Rahmen dieser Experimente keine signifikanten Unterschiede zwischen den A3C *BE* Komplexen nachgewiesen werden.

Darauffolgende Experimente fokussierten sich auf die Identifikation der Aminosäuren, die zu einer erhöhten *BE* Aktivität beitragen. Die *Loop1* und *αHelix1* Regionen der A3 Desaminasen spielen eine zentrale Rolle für die enzymatische Aktivität. Mittels PCR und Klonierung wurden weitere A3C *BE* Komplexe mit Mutationen dieser Regionen erstellt. Die Varianten wurden erneut dem fluoreszenz-basierten Echtzeit-Reporter-Assay unterzogen. Schlussendlich konnte ein *gain-of-function* des huA3C nach Erhalt der A3C CH2*αHelix1* und *Loop1* Aminosäuren Tyrosin und Glyzin an den Positionen 28 und 29 (^{28}YG^{29}) erzielt werden. Die detaillierte Charakterisierung von neuen *BE* Komplexen ist zentral für die Weiterentwicklung dieses vielversprechenden *gene editing* Tools.

Abstract

Human APOBEC3 (A3) represents a seven-member protein family of cytidine deaminases, A3A-D, F-H. The A3 enzymes are part of the innate immune system and represent restriction factors protecting the host from retroviral infection: through the deamination of cytidine (dC) to uridin (dU) residues within the viral genome, point mutations are inserted. Said mutations ultimately lead to the destabilization of the genetic material and thereby restrict virus replication.

Due to their catalytic activity as deaminases, A3 proteins are applied in the context of base editing (BE), which was derived from the CRISPR-Cas9 method. In contrast to the conventional CRISPR-Cas9, which requires DNA double-strand breaks, BE complexes introduce point mutations through deamination. Uracil glycosylase inhibitors (UGI) and enzymatically deactivated Cas9 (Cas9n) ensure that the point mutation is maintained. Similar to CRISPR-Cas9, BE requires a single guide RNA in order to recognize the DNA target.

Currently, of the A3 protein family, A3A, A3B and A3G are most commonly used for BE. Human A3C (huA3C) has so far shown low editing efficiency. A panel of six different base editors carrying A3 variants was created for this study. The A3A base editor served as a positive control. The remaining five BE complexes included huA3C and four A3C variants. Base editors were subsequently submitted to a fluorescence-based real-time reporter assay to test for BE activity. The aim was to identify an efficient A3C base editor and to characterize the A3C variant in its role as base editor. HEK293T cells were transiently transfected with the BE complexes, a sgRNA, and the reporter plasmid. BE activity could be observed both microscopically and through flow cytometry. While variant A3C Chimera 2 (A3C CH2) showed high efficiency, the remaining four A3C deaminases showed low to no BE efficiency. A3C CH2 is a chimera of huA3C with sooty mangabey A3C (smmA3C)-like. The first 37 amino acid residues match with smmA3C-like, whereas the remaining residues of the enzyme are derived from huA3C.

Various theories surged to explain why A3C CH2 in particular showed high editing frequencies, including deviant protein folding of the other variants, differences in intracellular localization, and a stronger affinity ofCas9 to A3C CH2. However, no significant differences were observed between the A3C base editors within these experiments.

Subsequent experiments focused on identifying the amino acid residues that contributed to A3C CH2s increased BE efficiency. *Loop1* and *αHelix1* regions of A3 deaminases are crucial for their enzymatic activity. Therefore, additional variants of huA3C and A3C CH2 with mutations in Loop1 and αHelix1 regions were created. The variants were then subjected to the fluorescence-based real-time reporter assay. A *gain-of-function* of huA3C was achieved after receiving the *αHelix1* of A3C CH2 and *Loop1* amino acid residues tyrosine and glycine at positions 28 and 29 (^{28}YG^{29}).

The detailed characterization of deaminases for BE contributes to ongoing efforts to improve this promising gene editing tool.

List of abbreviations

+ *positive*
°C *Degree Celsius*
µl *Microliter*
µM *Micromolar*
A3A *APOBEC3A*
A3B *APOBEC3B*
A3C CH2 *APOBEC3C Chimer 2*
A3F *APOBEC3F*
A3G *APOBEC3G*
A3H *APOBEC3H*
ABE *Adenine base editing*
ABE9 *Adenine base editor 9*
AIDS *Acquired immunodeficiency syndrome*
APOBEC3 Apolipoprotein B mRNA Editing Catalytic Polypeptide-Like 3
APS *Ammonium persulfate*
ART *Antiretroviral therapy*
BE *Base Editing*
bp *Base pair*
C *Cytosine*
CA *Capsid*
Cas9 *CRISPR-associated protein 9*
CBE *Cytidine base editing*
CCR5 C-C *chemokine receptor type 5*
CD4 *Cluster of differentiation 4*
CRISPR *Clustered Regularly Interspaced Short Palindromic Repeats*
CXCR4 *C-X-C chemokine receptor type 4*
dA *Adenine*
DAPI *4',6-diamidino-2-phenylindole*
dC *Cytidin*
ddH_2O *Double-distilled water*
dG *Guanine*
dI *Inosine*
DMEM *Dulbecco's High-Glucose Modified Eagle's Medium*
DMSO *Dimethyl sulfoxide*
DNA *Deoxyribonucleic acid*
DSB *Double strand breaks*
dsDNA *Double-strand desoxyribonucleic acid*
dT *Thymidine*
dTNPs *Nucleoside triphosphates*
dU *Uridin*
E *Glutamate*
E. coli *Escherichia coli*
ECL *Enhanced chemiluminescent luminol*
EDTA *Ethylenediaminetetraacetate*
eGFP *Enhanced green fluoresence protein*
env *Envelope*
Ext. *External*
FACS *Fluorescence-activated cell sorting*
FBS *Fetal bovine serum*
FITC *Fluorescein Isothiocyanate*
FP *Forward primer*
FSC *Forward scatter*
FSC-A *Forward Scatter-Area*
FSC-H *Foward Scatter- Height*
g *Relative centrifugal field*
G *Gylcine*
G2 *Second gap phase*
gag *Group specific antigen*
GFP *Green fluoresence protein*
gp120 *Surface glycoprotein*
gp160 *Precursor glycoprotein*
gp41 *Transmembrane protein*
H_2O *Water*
HA *Hemagglutinin*
HEK293T *Human embryonic kidney 293 T cells*
HIV *Human immunodeficiency virus*
HRP *Horseradish peroxidase*
huA3C *human APOBEC3C*
I *Isoleucine*
IgG *Immunoglobulin G*
IN *Integrase*
Int. *Internal*
IP *Immunoprecipitation*
K *Lysine*
kb *Kilobyte*
l *Liter*
LB *Lysogeny broth*

LINEs *Long interspersed nuclear elements*
LTR *Long terminal repeat*, *Long terminal repeats*
M *Molar*
MA *Matrix*
mCherry *monomeric Cherry*
mg *Milligram*
mM *Millimolar*
mRNA *messenger ribonucleic acid*, *messenger ribonucleic acid*
NaCl *Sodium chloride*
NaOH *Sodium hydroxide*
NC *p7*
NEF *Negative regulatory factor*
ng *Nanogram*
NLS *Nuclear localization sequence*
nm *Nanometer*
NS *Non-specific*
P *Probability*
p1 *Spacing peptide 1*
p17 *Matrix*
p2 *Spacing peptide 2*
p24 Capsid protein
p6 *Budding-related peptide*
p7 *Nucleocapsid*
PAGE *Polyacrylamide gel electrophoresis*
PBS *Phosphate-buffered saline*
PCR *Polymerase chain reaction*
PE *Phycoerythrin*
pH *Potential of hydrogen*
pol *Polymerase*
PR *Protease*
PVDF *Polyvinylidene fluoride*
R *Arginine*
rAPOBEC1 *rat APOBEC1*
REV *Regulator of expression of virion proteins*
RIPA *Radioimmunoprecipitation assay*
RP *Reverse primer*
rpm *Rounds per minute*
RRE *rev response element*
RT *Reverse transcriptase*
S *Serine*
SD *Standard deviation*
SDS *Sodium dodecyl sulfate*
sgRNA *single guide ribonucleic acid*
SIV Δ*vif* *viral infectivity factor-deficient Simian immunodeficiency virus*
SIVagm *Simian Immunodeficiency Virus from the African green monkey*
smmA3C *sooty mangabey APOBEC3C*
SSC *Side scatter*
SSC-A *Side Scatter-Area*
ssDNA *Single-stranded deoxyribonucleic acid*
SU *Surface unit*
TAE *Tris-Acetate-EDTA*
TAT *Trans-activator of transcription*
TBST *Tris-buffered Saline with Tween 20*
TE *Tris-EDTA buffer*
TMED *Tetramethylethylenediamine*
Tris *Tris(hydroxymethyl)aminomethane*
Tris-HCl *Tris Hydrochloride*
TU *Transmembrane unit*
U *Unit*, *Uracil*
UDG *Uracil DNA glycosylase*
USA *United States of America*
UV *Ultraviolet*
V *Volt*
V5 *V protein of the simian virus 5*
VIF *Viral infectivity factor*
VLPs *Viral-like particles*
VPR *Viral protein r*
VPU *Viral protein u*
VSV-G *Vesicular stomatitis virus G protein*
W *Tryptophan*
Y *Tyrosine*
α *Anti*
αHelix1 *Alpha Helix 1*
Δ*VIF* *viral infectivity factor-deficient*
ψ *Psi*

Table of contents

1. Introduction

1.1. Human immunodeficiency virus and acquired immune deficiency syndrome

The Human Immunodeficiency Virus (HIV) is a significant global health concern, causing chronic immunodeficiency and leaving HIV-positive individuals vulnerable to a wide range of opportunistic infections and cancers. Without effective treatment, HIV can lead to the acquired immunodeficiency syndrome (AIDS), the final stage of infection associated with a low CD4+ T cell count and specific AIDS-defining diseases (Ford et al., 2025). As of today, over 39 million people worldwide are living with HIV and in spite of significant advances in treatment, there is still no cure available (UNAIDS, 2024; World Health Organization, 2024).

HIV is transmitted through blood, semen, vaginal fluids, rectal fluids, and breast milk. The most common routes of transmission are unprotected sexual contact, sharing of needles, and mother-to-child transmission during childbirth or breastfeeding. HIV transmission can be minimized through prevention strategies such as safe sex practices, needle exchange programs, and antiretroviral therapy (ART) (Ghosn et al., 2018). HIV-positive individuals with an undetectable virus load cannot transmit HIV sexually (Okamoto et al., 2024). Furthermore, the effective treatment reduces the likelihood of transmission during pregnancy and childbirth significantly (Poliektov & Badell, 2023).

HIV infection progresses through several stages. Following the initial exposure, the virus rapidly replicates leading to an acute HIV infection. During this stage individuals, may present flu-like symptoms over a period of days to weeks. An acute HIV infection, however, may also be completely asymptomatic. If untreated, the virus enters a chronic phase, gradually depleting CD4+ T cells and continuously impairing the immune system. This phase is typically asymptomatic and can last multiple years. Eventually, HIV-positive individuals become susceptible to opportunistic infections. Without treatment, they develop AIDS-defining diseases, such as *Pneumocystis-jirovecii* pneumonia, Kaposi sarcoma or cerebral toxoplasmosis (Attia et al., 2025; World Health Organization, 2024).

ART allows for an efficient treatment through the suppression of viral replication. In spite of great advancements in the field of ART, its effectiveness still greatly depends on the

regular uptake of medication. As HIV integrates its genome into durable cells like memory cells, interrupting treatment can cause a rebound and a subsequent increase in viral load. This virus reservoir poses a great challenge for a potential cure of HIV. Ongoing research explores amongst others approaches such as gene editing technologies and immune-based therapies to target and eliminate these reservoirs (Gurrola et al., 2024; Kulpa et al., 2025). Additionally, ensuring consistent treatment adherence and expanding global access to ART remain crucial for controlling HIV (Cowan et al., 2025; Ford et al., 2025; Ghosn et al., 2018).

1.2. HIV genome and structure

HIV belongs to the *Lentivirus* genus of the *Retroviridae* family. There are two main types of HIV: HIV-1 and HIV-2. HIV-1 is responsible for the vast majority of global infections, while HIV-2 is less common and is primarily found in West Africa. Both subtypes share similar mechanisms of infection and replication. In addition to the two main subtypes, HIV-1 is further divided into groups and subtypes reflecting the virus's genetic diversity (Sharp & Hahn, 2011).

The ~9 kb long proviral genome of all retroviruses consists of three primary genes: group-specific antigen (*gag*), polymerase (*pol*) and envelope (*env*). Both ends of the genome are flanked by long terminal repeats (LTR) (see Fig.1). *Gag* encodes the precursor polyprotein p55Gag, which is cleaved the matrix (p17), capsid protein (p24), nucleocapsid (p7), the budding-related peptide p6, and two spacer peptides (p1 and p2). Together they form the viral core. The viral enzymes reverse transcriptase (RT), integrase (IN), and protease (PR) are encoded by pol. The enzymes are crucial for viral replication, genome integration, and protein processing. *Env* encodes gp160, a precursor glycoprotein that is processed into the surface glycoprotein gp120 and the transmembrane protein gp41. These proteins mediate viral entry by binding to CD4 and co-receptors (CCR5 and CXCR4). Cleaving processes take place in the endoplasmic reticulum and Golgi apparatus. Finally, the LTR are essential for viral transcription, integration and gene expression (Kwong et al., 1998; Lesbats et al., 2016; Schemelev et al., 2024).

In addition to these structural proteins, lentiviruses also carry genes for regulatory and accessory proteins. In the case of HIV-1, these are TAT and REV and VPR, VIF, VPU, and NEF. As a regulatory protein, the trans-activator of transcription (TAT) increases the efficiency of viral RNA transcription. The regulator of expression of virion proteins (REV) is essential for the translation of viral proteins: REV escorts unspliced and partially spliced mRNA from the nucleus to the cytoplasm. The viral protein r (VPR) is crucial for efficient viral replication and genome integration. VPR is involved in the nuclear import of the preintegration complex, induces a G2-phase cell cycle arrest, and enhances LTR activity. The accessory proteins viral infectivity factor (VIF), viral protein u (VPU), and negative regulatory factor (NEF) primarily allow the virus to evade the host immune system and counteract host restriction factors. A more detailed characterization of VIF and its functions is given below. All three proteins are able to interact with and through this interaction downregulate or restrict host proteins. Structural, regulatory, and accessory proteins are expressed through both full-length and spliced RNA (Karn & Stoltzfus, 2012; Rashid et al., 2024; Schemelev et al., 2024). A schematic illustration of the HIV-1 genome is given in Fig. 1.

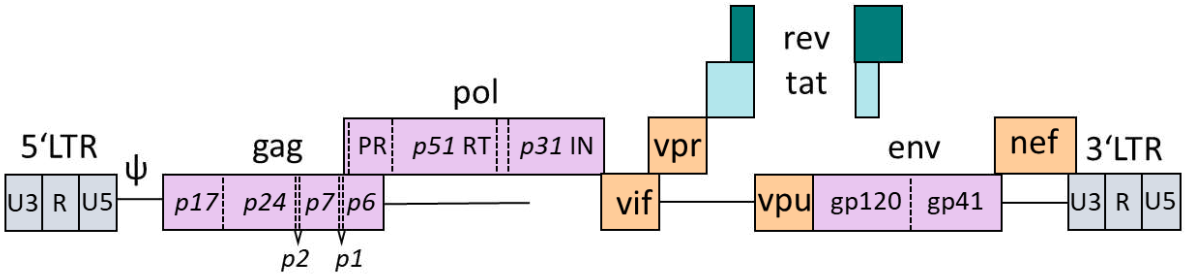

Figure 1. Graphical illustration of the HIV-1 genome. In addition to the three structural, primary genes, *gag*, *pol*, and *env*, regulatory *(rev and tat*) and genes for accessory proteins (*nef, vpr, vpu, and vif*) are displayed. The genome is flanked by LTRs at both the 5'- and 3'-terminal ends. The packaging signal Psi (ψ) is located downstream of the regulatory elements of the LTRs, namely U3, R, and U5. Core structural proteins are encoded by *gag* and include matrix (MA, p17), capsid (CA, p24), and nucleocapsid (NC, p7, p6) proteins. The viral enzymes protease (PR), reverse transcriptase (RT, 51), and integrase (IN, p31) are encoded by the *pol* gene. The regulatory genes and accessory proteins, VIF, VPR and VPU are located downstream of the *pol* gene. The viral envelope is encoded by the *env* and consists of the surface unit (SU, gp120) and transmembrane unit (TU, gp41). The fourth accessory protein, NEF, follows the *env* gene.

1.3. Deaminases as retroviral restriction factors

The APOBEC3 (A3) protein family is a critical component of the human innate immune system and consists of seven members: A3A-A3D and A3F-A3H. These proteins belong to a larger family of cytidine deaminases that catalyze the deamination of cytosine (C) to uracil (U) in single-stranded DNA (ssDNA) (Fig. 2) (Salter et al., 2016). Due to their enzymatic activity, the A3 proteins represent restriction factors of the host innate immune system in defense against retroviral infections, including HIV (Aguiar et al., 2008; Goila-Gaur & Strebel, 2008; Harris et al., 2003).

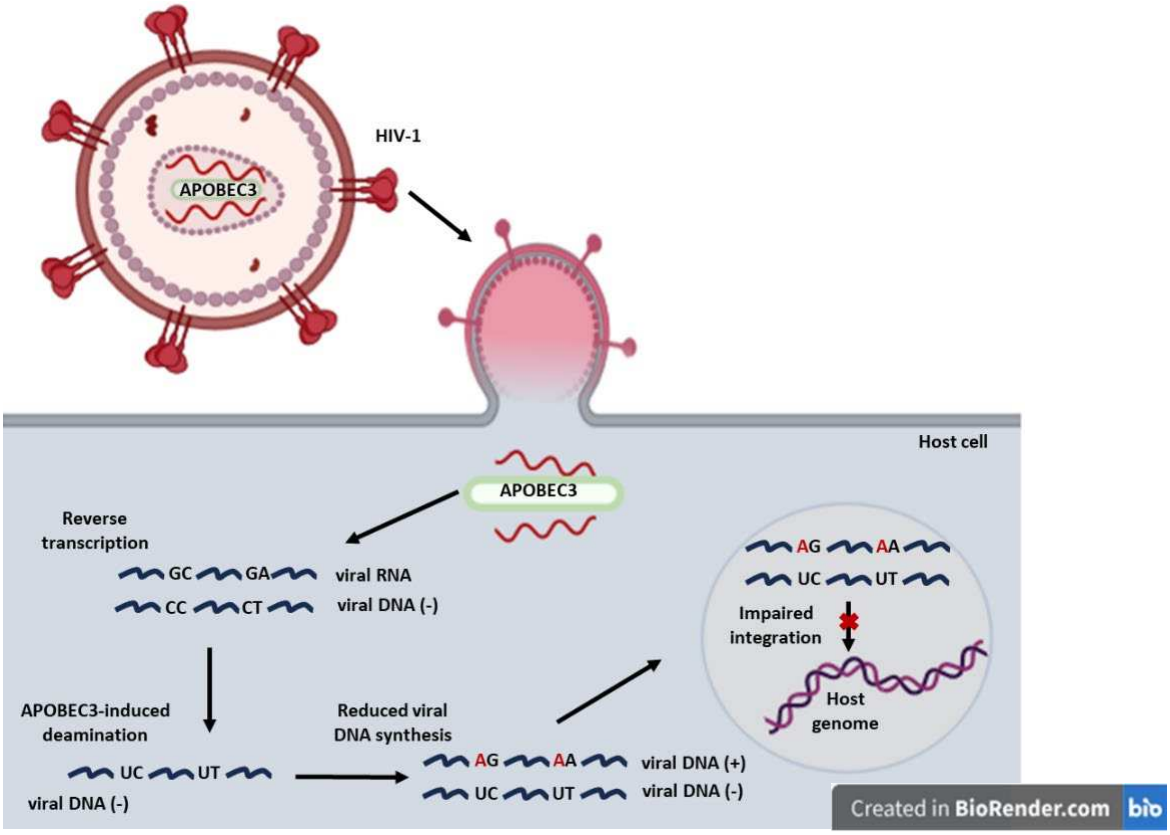

Figure 2. APOBEC3 (A3) mechanism. Graphical illustration of A3-induced hypermutations of the viral genome. Once HIV-1 enters the cell, the viral genome (RNA) is released and transcribed into a negative sense (-) strand viral DNA. After reverse transcription, APOBEC3 introduces dC-to-dT conversions to this single strand DNA. These mutations lead to a reduced viral DNA synthesis with mutations to the positive sense (+) strand DNA. This ultimately results in an impaired integration of the viral DNA in the host genome.

A3 enzymes exert their antiviral effects by targeting the viral genomes during replication. They specifically recognize and deaminate ssDNA desoxy-cytosine (dC) residues within the viral genome, converting the targeted residue into desoxy-uridine (dU). This deamination leads to a mismatch during viral DNA replication, where the opposing DNA

strand incorporates desoxy-adenine (dA) opposite the dU. In subsequent rounds of replication, the dU is recognized as desoxy-thymidine (dT), resulting in a cytosine-to-thymine (dC→dT) conversion. These base modifications in the minus-strand DNA lead to coding changes and the introduction of premature stop codons in the plus-strand viral genome (dG to dA hypermutation). Accumulating point mutations lead to a destabilization of the viral genetic material, which in turn severely impairs its ability to replicate effectively (Fig. 2) (Bishop et al., 2004; Harris et al., 2003; Shi et al., 2017). Beyond their mutagenic activity, virus-incorporated A3 enzymes also exhibit deaminase-independent antiviral mechanisms, including interference with reverse transcription and inhibition of viral DNA integration. APOBEC3 enzymes therefore play a crucial role in limiting viral replication, controlling viral load, and delaying retroviruses from spreading in infected hosts (Jaguva Vasudevan et al., 2020).

Lentiviruses, however, have developed countermeasures in response to the antiviral activity of A3 enzymes. As mentioned above, HIV encodes for the accessory protein VIF, which serves as a key antagonist to A3 enzymes (Fig. 3). VIF is able to interact with and bind A3 proteins, targeting them for ubiquitination. This subsequently leads to the proteasomal degradation of A3 enzymes through the host's own cellular machinery. This process prevents A3 enzyme incorporation in newly assembling viral particles leading to infective virions (Fig. 3). Through VIF, HIV is capable of escaping this critical arm of the innate immune defense (Kitamura et al., 2012).

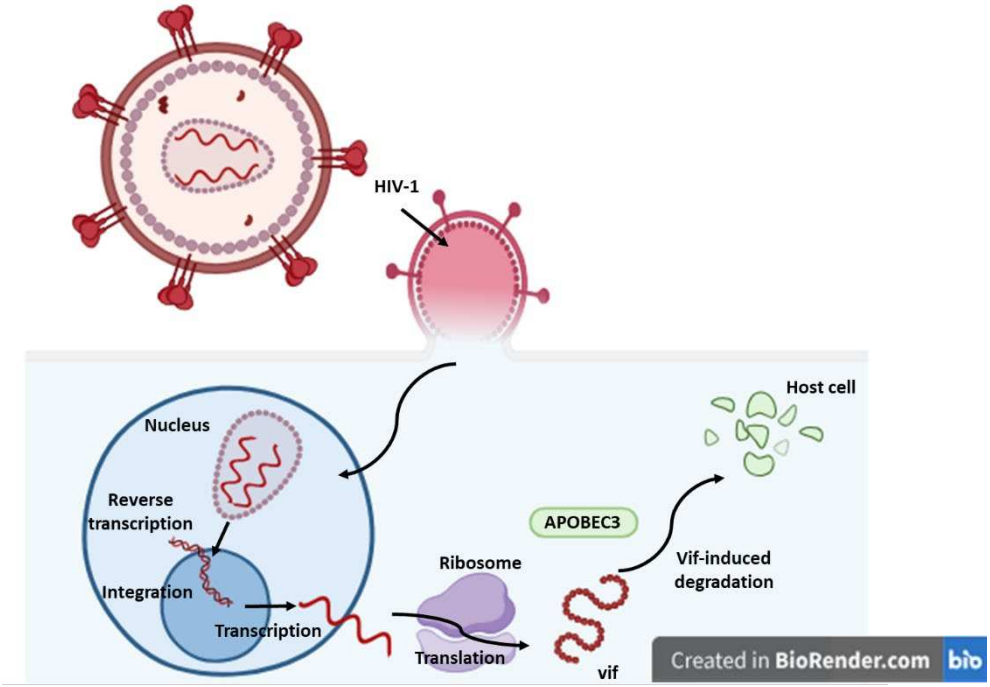

Figure 3. HIV-1 accessory protein viral infectivity factor (VIF). Graphical representation of the lentiviral counter-mechanism. After infection, the HIV-1 genome is transcribed in the nucleus and integrated into the host genome. After integration the host replication machinery transcribes the viral genes. The newly transcribed RNA is then translated in the host ribosomes. One of the translated proteins is the accessory protein VIF. VIF induces the degradation of A3 enzymes, circumventing hypermutations and the encapsidation of A3 enzymes in newly assembling virions.

Despite this viral evasion strategy, APOBEC3 enzymes remain significant contributors to the human antiviral response. Their activity is not limited to retroviruses; they also play roles in restricting other viral families and endogenous retroelements, such as long interspersed nuclear elements (LINEs) (Biolatti et al., 2018; Jaguva Vasudevan et al., 2020; Modenini et al., 2022).

Interestingly, the activity of A3 enzymes extends beyond viral defense and has relevance in other biological contexts. For example, their mutagenic potential is associated with certain cancer types, where they contribute to the somatic hypermutation of DNA, leading to tumor progression. For instance, in breast cancer both A3A and A3B overexpression has been observed and is considered a significant source of mutations associated with these tumors. This dual role as both protectors against viral infections and contributors to genomic instability highlights the complex nature of A3 enzymes making them a double-edged sword (Dennis et al., 2024; C. Zhang et al., 2024).

In the context of HIV infection, the interplay between A3 enzymes and the VIF protein remains a focal point of research. By disrupting VIF-mediated degradation of A3 proteins, it may be possible to restore their mutagenic activity and reduce HIV replication, providing a novel strategy for antiretroviral therapy (Rashid et al., 2024).

1.4. The CRISPR/Cas9 derived tool of base editing

Base editing (BE) is a precise genome editing method based on modified CRISPR/Cas9 systems. A fundamental advantage of BE over the conventional CRISPR/Cas9 method is that BE does not require potentially harmful double strand breaks (DBS) (Chen et al., 2024; Komor et al., 2016). BE allows for a selective exchange of individual bases in a ssDNA strand of the chromosomal DNA. This enables a more controlled and accurate

editing process whilst avoiding the risk of unintended genomic rearrangements or large-scale mutations often associated with DSBs (Naeem et al., 2020).

Two forms of BE are distinguished: Adenine Base Editing (ABE) and Cytidine Base Editing (CBE). In ABE, an adenosine deaminase is used to convert adenine (dA) into inosine (dI). Inosine is then read as guanine (dG) during DNA replication, resulting in a precise adenine-to-guanine (dA→dG) mutation (Jeong et al., 2021) (Fig. 4a). On the other hand, CBE requires a cytosine deaminase to convert cytosine (dC) into uracil (dU). During DNA replication dU is then interpreted as thymine (dT), resulting in a dC→dT conversion. Normally, however, during DNA replication and repair, dU is recognized and usually removed by a cellular enzyme called uracil DNA glycosylase (UDG). In order to prevent the excision of the dU residue, CBE complexes carry an uracil glycosylase inhibitor (UGI) which is able to bind UDG (Fig. 4b)

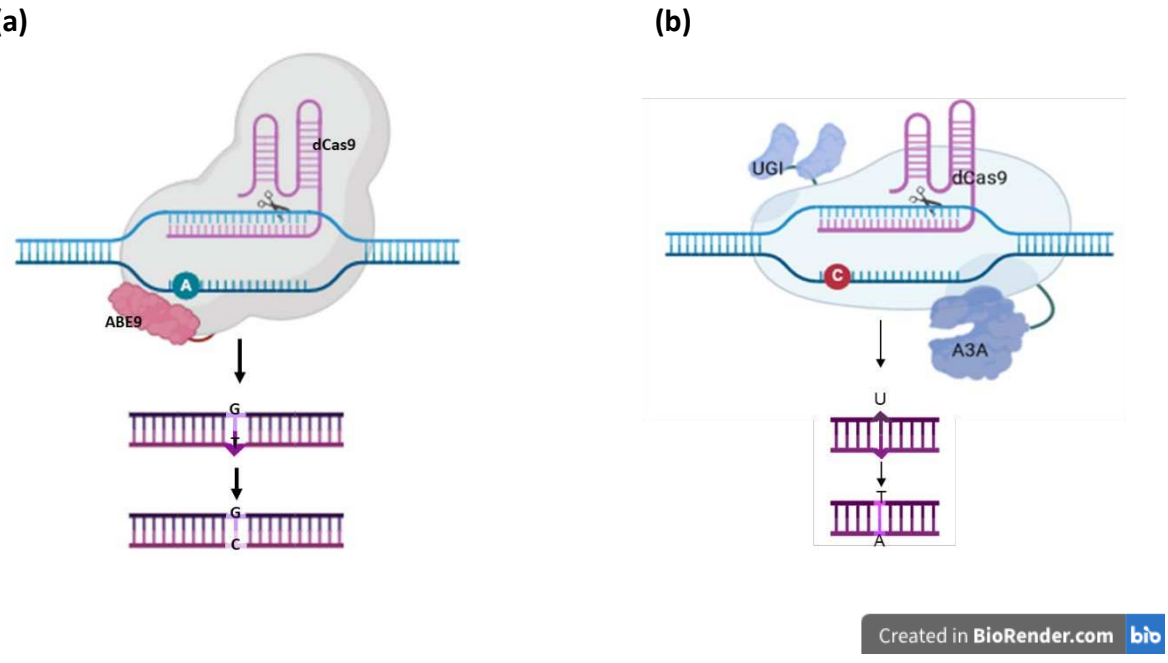

Figure 4. Graphical illustration of base editing. (a) Adenine base editors induce the deamination of a specific dA target. The established ABE9 desaminase was chosen as a representative for this graphical depiction. ABE9 was chosen as a representative base editor and allows for a precise dA-to dG conversion. The editing complex furthermore carries an enzymatically deactivated Cas9 protein, which promotes the fixation of the induced point mutation. A singe guide RNA (sgRNA) allows for target recognition. (b) Cytidine base editor-induced deamination of a specific dC target. A3A was chosen as representative cytidine deaminase, the construct additionally carries an uracil glycosylase inhibitor (UGI), an enzymatically deactivated Cas9 protein, and is guided by a sgRNA. The editing complex enables a dC-to-dU conversion, which finally leads to a dC-to-dT mutation.

In addition to the deaminases, BE complexes include several other crucial components. Similar to the CRISPR/Cas9 method, BE requires a single-stranded guide RNA (sgRNA). The sgRNA is required for target recognition and directs the BE complex to the desired location within the DNA by binding to the target sequence. Target binding requires a Cas9 enzyme or a variant thereof. BE employs Cas9 nickase (nCas9). Once nCas9 binds to the genomic target sequence, the editing window is exposed allowing the deaminase to modify the targeted base. nCas9 creates a nick within the non-edited DNA strand opposite to the uracil introduced by the deaminase. This activates cellular repair mechanisms and stimulates the resynthesis of the non-edited strand. The edited strand serves as template. This mechanism is essential for BE as it results in the desired modification of both DNA strands enabling a precise base pair exchange (Komor et al., 2017; Trevino & Zhang, 2014).

Base editors exhibit a wide arrange of editing abilities making BE a highly versatile tool applicable in numerous contexts. Within medical areas, BE shows great promise for the treatment of genetic disorders caused by point mutations. Preclinical trials have shown encouraging results for the application of BE to treat sickle cell anemia, beta-thalassemia, and Duchenne muscular dystrophy. When compared to conventional CRISPR/Cas9, BE is a safer alternative as it does not rely on DSB reducing the risk of large-scale genomic damage or undesirable genetic rearrangements (Porto et al., 2020; Rees & Liu, 2018). In addition to directly treating genetic diseases through BE, emerging strategies suggest the application of BE to modify protein epitopes, enabling the creation of shared antigen targets for blood cancer immunotherapy (Wellhausen et al., 2023).

Aside from its application in medical fields, BE is also used to apply genetic modifications in plants. BE enables trait improvements: single base exchanges can lead to higher drought tolerance, nutrient content, and disease resistance (Mishra et al., 2020).

However, despite its precision, BE does face several challenges. Although off-target mutation rates are lower compared to conventional CRISPR/Cas9, BE can still lead to undesired mutations outside the editing window. The nCas9 enzyme defines the editing window, which typically has a width of three to nine nucleotides. Within this editing window the base editors may be able to editing various bases. Further narrowing the editing window is therefore a fundamental approach to circumvent these potential

unwanted editing events and increase precision. Another aspect that has to be considered is the deaminases sequence bias. APOBEC1 for instance has a strong editing preference for TC sequences over GC. Creating new base editors with deaminases without a strong sequence bias broadens the spectrum of the application of the editing constructs. These mentioned efforts to improve BE additionally reduces off-target effects (Y. Zhang et al., 2024; Zhao et al., 2024).

The efficient delivery and proper introduction of BE constructs into target cells are crucial steps successful *in vivo* therapeutic applications. Two main delivery methods are distinguished and are base either on viral or non-viral systems. Viral vectors, such as adeno-associated viruses, adenoviruses, and lentiviruses, are widely used due to their ability to reach target cells with high efficiency. Furthermore, the viral vectors allow for a stable base editor expression. On the other hand, non-viral delivery strategies, such as lipid nanoparticles and virus-like particles (VLPs), enable transient base editor expression. The transient protein expression may reduce off-target effects and potential health safety concerns (Newby & Liu, 2021; Rees et al., 2017; Zhang et al., 2022; Zhou et al., 2022).

Another challenge BE faces is limited efficiency with low on-target edits depending on the genomic context. Factors such as DNA accessibility, chromatin structure, and the presence of competing repair pathways can impact editing efficiency (Huang et al., 2021; Liu et al., 2020).

In order to address these issues, ongoing research is conducted to continuously refine this editing technique. Efforts for improvement include applying both wild type and modified deaminases base editors. Additionally, strategies to optimize the design of sgRNAs and minimize off-target effects are also being explored. The developing of tools that predict potential off-targets can be used for an improved sgRNA design (Hu et al., 2021; Zhang et al., 2023). A further component subjected to improvement is the nCas9: novel Cas9 variants with greater precision are being developed (Slesarenko et al., 2022).

The ongoing efforts to improve BE efficiency make the technique a highly promising editing tool (Rees & Liu, 2018).

1.5. The role of APOBEC enzymes as part of CBE complexes

As mentioned previously, CBEs enable precise dC-to-dT base conversions due to a cytidine deaminase as part of the BE complex. One of the most critical factors in the development of CBEs is the choice of the deaminase, which greatly determines editing efficiency. The APOBEC family has been widely explored for use in CBEs due to its inherent ability to catalyze dC-to-dU conversions in DNA (Jin et al., 2020; Liu et al., 2020; Martin et al., 2019).

One of the first cytidine deaminases used in CBEs was rat APOBEC1 (rAPOBEC1), a deaminase derived from the RNA-editing APOBEC1 protein in rats. Rat (r) APOBEC1 demonstrated the ability to deaminate cytosine in ssDNA, making it a suitable candidate for early BE constructs. However, the broad editing window of rAPOBEC1 lead to a high frequency of off-target events. Additionally, rAPOBEC1 displays a high sequence preference , limiting its utility to very specific DNA sequences (Liu et al., 2023).

To broaden the spectrum of potential applications, rAPOBEC1 was replaced by human APOBEC3 deaminases. Among these, APOBEC3A (A3A), APOBEC3B (A3B), and APOBEC3G (A3G) have demonstrated the highest cytosine editing frequencies *in vitro* (Gehrke et al., 2018; Jin et al., 2020; Liu et al., 2020; Ren et al., 2024).

The superior performance of A3A, A3B, and A3G-based constructs is attributed to several factors. A3A, for instance, exhibits high catalytic activity and a broad sequence compatibility. It can efficiently deaminate cytosines in various sequence contexts, providing a versatile editing tool for challenging loci. A3B and A3G also demonstrate strong performance due to their stability, which enhances their activity within the editing complex. These deaminases have been shown to maintain robust activity even in complex genomic environments, ensuring high editing efficiency across a wide range of applications (Jin et al., 2020; Langlois et al., 2005).

Conversely, CBEs that incorporate other A3 family members, such as APOBEC3C (A3C) and APOBEC3F (A3F), exhibit markedly reduced editing efficiencies. A3C has shown little to no BE activity. This may be attributed to a narrower editing window and high sequence preference, limiting utility for targets outside the preferred context. A3F both exhibits reduced catalytic efficiency and potentially lower expression stability, leading to fewer

editing events. These challenges have made A3C and A3F over all less favorable choices for most base editing applications (Langlois et al., 2005; Martin et al., 2019).

1.6. Human A3C and variants with high catalytic and antiviral activity

Human A3C (huA3C) as part of the APOBEC family is part of the innate immune system in response to retroviral infection. While huA3C shows high antiviral activity against VIF-deficient Simian immunodeficiency virus (SIV Δ*vif*), its effectivity against HIV-1 Δ*VIF* is very limited. Interestingly, A3C does inhibit the infectivity of herpes simplex virus, some human papilloma virus strains, hepatitis B virus, and others. Some studies have focused on gaining a deeper understanding of the antiviral activity of huA3C (Jaguva Vasudevan et al., 2020).

The A3C S188I variant is a point mutation in the huA3C gene, where serine (S) at position 188 is substituted by isoleucine (I). This single nucleotide polymorphism is associated with an increased capability to inhibit HIV-1 (Wittkopp et al., 2016).

Efforts to further identify determinants of A3 antiviral effectivity have revealed that the enzyme's loop1 region is highly important (Ziegler et al., 2019). Substituting two amino acid residues within huA3C loop 1—replacing tryptophan and glutamate (WE) with arginine and lysine (RK)—greatly improves A3C's ability to interact with ssDNA. This huA3C variant named A3C WE.RK displays enhanced deaminase activity (Jaguva Vasudevan et al., 2020). The creation of chimera of huA3C combined with sooty mangabey monkey A3C-like (smmA3C-like) helped to further identify and characterize highly effective huA3C variants. smmA3C-like itself is an A3C-like protein derived from a synthetic open reading frame from sooty the mangabey monkey. smmA3C-like is a potent deaminase able of inhibiting HIV-1 at a similar level as human A3G (Jaguva Vasudevan et al., 2020).

Out of the chimeras, especially A3C Chimer2 (A3C CH2) proved to be a highly efficient inhibitor of HIV-1 Δ*vif* replication. In order to create A3C CH2, the N-terminal 37 amino acids of huA3C were replaced by smmA3C-like's residues. These 37 residues harbor αhelix 1 and loop1 regions which are both crucial for enzymatic activity (Jaguva Vasudevan et al., 2020).

1.7. Objectives

Based on the findings presented above, the hypothesis surged that the huA3C variants with increased enzymatic and antiviral activity would also show BE efficiency. We thus created a set of six base editors: huA3C, A3C WE.RK, A3C S188I, A3C CH2, and smmA3C-like were incorporated into BE complexes. Proven efficiently as part of a BE complex, A3Ai, fused to nCas9 via a L1 intron, was used as positive control. To assess BE activity, we co-expressed A3 variants fused to nCas9-UGI, guided by a sgRNA, along with a reporter system that produced a green fluorescent signal upon CBE.

After successfully singling out A3C CH2 as efficient deaminase for BE, we set out to further characterize the enzyme in the context of BE. Experiments were focused on identifying determinants for BE efficiency.

2. Material and methods

2.1. Material

2.1.1. Electronic equipment

Device	**Producer**
Cell sorter FACSAria III	BD Biosciences, Heidelberg, Germany
Centrifuge 4K15	Sigma Laborzentrifugen GmbH, Osterode am Harz, Germany
Centrifuge Heraeus Fresco 21 (cell culture, lysates)	Thermo Fisher Scientific, Waltham, USA
Centrifuge Heraeus Heraeus Pico 21 (cell culture, lysates)	Thermo Fisher Scientific, Waltham, USA
CO_2 Incubator	IBS, Fernwald, Germany
Confocal microscope LSM 510 Meta	Carl Zeiss, Cologne, Germany
Film development apparatus, Cruix 60	AGFA Healthcare, Mortsel, Belgium
Gel documentation system/ UV transilluminator	Peqlab, Erlangen, Germany
Heating block	Fisher Biobloc Scientific, Illkirch Cedex, France
Incubation shaker Thriller	Peqlab, Erlangen, Germany
Incubator, KS 4000 I Control	IKA, Staufen, Germany
Microscope AE29	Motic, Barcelona, Spain
Microscope Cell Observer Z1	Carl Zeiss, Cologne, Germany
Nano-Drop NP-1000	Peqlab, Erlangen, Germany
PowerPac Basic Power Supply	BioRad, Munich, Germany
PowerPac HC	BioRad, Munich, Germany
Shaker type 3006	GFL, Brugwedel, Germany
Thermocycler GeneTouch	Bioer Technology, Hangzhou, China

Thermocycler T3	Biometra Biomedizinische Analytik, Jena, Germany
ThermoMixer Compact	Eppendorf, Hamburg, Germany
TransBlot SD Semi-Dry Transfer Cell	BioRad, Munich, Germany
Vortex 2	IKA, Staufen, Germany

Table 1: Electronic equipment

2.1.2. Chemical reagents

Chemical reagent	**Producer**
4',6-diamidino-2-phenylindole (DAPI)	Thermo Fisher Scientific, Waltham, USA
Acrylamid (30%)/ Bisacrylamid (0,8%)	Applichem, Darmstadt, Germany
Agar	Applichem, Darmstadt, Germany
Agarose	Bio & Sell GmbH, Feucht, Germany
Ammonium persulfate (APS)	Carl Roth, Karlsruhe, Germany
Ampicillin	Sigma-Aldrich, Taufkirchen, Germany
anti-HA affinity matrix beads	Roche, Basel, Switzerland
Caseinpepton	Carl Roth, Karlsruhe, Germany
DNA loading buffer (6x)	Thermo Fisher Scientific, Waltham, USA
Ethanol	Applichem, Darmstadt, Germany
Ethidium bromide solution, 0,025%	Carl Roth, Karlsruhe, Germany
Ethylenediaminetetraacetate (EDTA)	Carl Roth, Karlsruhe, Germany
Glycerol	Applichem, Darmstadt, Germany
Isopropanol	Carl Roth, Karlsruhe, Germany

Marker PageRuler Prestained Protein Ladder	Thermo Fisher Scientific, Waltham, USA
Methanol	Applichem, Darmstadt, Germany
Milk powder	Applichem, Darmstadt, Germany
NonidetP-40 (NP-40)	Applichem, Darmstadt, Germany
Phenylmethanesulfonyl fluoride solution	Sigma-Aldrich, Taufkirchen, Germany
Proteaseinhibitor Cocktail Set III	Calbiochem, Darmstadt, Germany
Sodium chloride (NaCl)	MP Biomedicals, Santa Ana, USA
Sodium deoxycholate/ deoxycholic acid	Carl Roth, Karlsruhe, Germany
Sodium dodecyl sulfate (SDS)	Applichem, Darmstadt, Germany
Sodium hydroxide (NaOH)	Applichem, Darmstadt, Germany
Tetramethylethylenediamine (TMED)	MP Biomedicals, Santa Ana, USA
Tris(hydroxymethyl)aminomethane (Tris)	Applichem, Darmstadt, Germany
Tween 20	Sigma-Aldrich, Taufkirchen, Germany
Yeast extract	Carl Roth, Karlsruhe, Germany

Table 2: Chemical reagents

2.1.3. Reagents for PCR and cloning

BsmBI restriction enzyme	Thermo Fisher Scientific, Waltham, USA
Buffer R	Thermo Fisher Scientific, Waltham, USA
Buffer Tango	Thermo Fisher Scientific, Waltham, USA
Cfr9I (XmaI) buffer	Thermo Fisher Scientific, Waltham, USA
Cfr9I (XmaI) restriction enzyme	Thermo Fisher Scientific, Waltham, USA
dTNPs (10 mM)	Thermo Fisher Scientific, Waltham, USA
KpnI buffer	Thermo Fisher Scientific, Waltham, USA
KpnI restriction enzyme	Thermo Fisher Scientific, Waltham, USA
NotI restriction enzyme	Thermo Fisher Scientific, Waltham, USA
Nuclease-free water	Qiagen, Hilden, Germany
Q5 DNA-polymerase	New England Biolabs, Ipswich, USA
Q5 DNA-polymerase buffer	New England Biolabs, Ipswich, USA
T4 DNA ligase	New England Biolabs, Ipswich, USA
T4 DNA Ligase buffer (10X)	New England Biolabs, Ipswich, USA
Tris-EDTA buffer (TE) buffer	In-house produced reagent
XhoI restriction enzyme	Thermo Fisher Scientific, Waltham, USA

Table 3: Reagents used for plasmid production

2.1.4. Plasmids

Plasmid	**Description**	**Origin/ Producer**
A3Ai-Cas9n-UGI-NLS	Base editing complex carrying genes for A3Ai, Cas9, UGI, and NLS	Addgene #109425
A3A-HA	A3A cloned in pcDNA3.1, with a C-terminal HA-tag	Bulliard, 2010
A3C CH2-Cas9n-UGI-NLS	Base editing complex carrying genes for A3C CH2, Cas9, UGI, and NLS	Produced in this study
A3C CH2.αHelix1$_{hu}$.	Base editing complex carrying genes for A3C CH2.αHelix1$_{hu}$, Cas9, UGI, and NLS	Produced in this study
A3C S188I-Cas9n-UGI-NLS	Base editing complex carrying genes for A3C S188I, Cas9, UGI, and NLS	Produced in this study
A3C WE.RK-Cas9n-UGI-NLS	Base editing complex carrying genes for A3C WE.RK, Cas9, UGI, and NLS	Produced in this study
A3C CH2-HA	A3C CH2 cloned in pcDNA3.1, with a C-terminal HA-tag	Jaguva Vasudevan, 2020
A3C S188I-HA	A3C S188I cloned in pcDNA3.1, with a C-terminal HA-tag	Jaguva Vasudevan, 2020
A3C WE.RK-HA	A3C WE.RK cloned in pcDNA3.1, with a C-terminal HA-tag	Jaguva Vasudevan, 2020

huA3C-Cas9n-UGI-NLS	Base editing complex carrying genes for huA3C, Cas9, UGI, and NLS	Produced in this study
huA3C.αHelix1$_{CH2}$-Cas9n-UGI-NLS	Base editing complex carrying genes for huA3C.αHelix1$_{CH2}$, Cas9, UGI, and NLS	Produced in this study
huA3C.αHelix1$_{CH2}$RK-Cas9n-UGI-NLS	Base editing complex carrying genes for huA3C.αHelix1$_{CH2}$RK, Cas9, UGI, and NLS	Produced in this study
huA3C-HA	huA3C cloned in pcDNA3.1, with a C-terminal HA-tag	Jaguva Vasudevan, 2020
L138S sgRNA	Single guide RNA, MLM3636 vector	Backbone Addgene #43860; insert created in this study
L202S sgRNA	Single guide RNA, MLM3636 vector	Backbone Addgene #43860; insert created in this study
NS sgRNA	Single guide RNA, MLM3636 vector	Backbone Addgene #43860; insert created in this study
plenti-CMV-mCherry-T2A-GFP L138S	Reporter plasmid, carrying genes for eGFP, mCherry, LTR	Backbone Addgene #109427
plenti-CMV-mCherry-T2A-GFP L202S	Reporter plasmid, carrying genes for eGFP, mCherry, LTR	Backbone Addgene #109427
pMD2.G	VSV-G packaging protein	Addgene #12259

psPAX2	Second generation lentival packaging plasmid, carries HIV-1 group M gag, pol, rev, tat, RRE	Addgene #12260
smmA3C-like-Cas9n-UGI-NLS	Base editing complex carrying genes for smmA3C-like, Cas9, UGI, and NLS	Produced in this study
smmA3C-like-HA	smmA3C-like cloned in pcDNA3.1, with a C-terminal HA-tag	Jaguva Vasudevan, 2020

Table 4: Plasmids used in this study

2.1.5. Culture media and reagents for cell culture

Medium/ reagent	**Producer**
Dimethyl sulfoxide (DMSO)	Merck Millipore, Darmstadt, Germany
Dulbecco's High-Glucose Modified Eagle's Medium (DMEM)	PAN-Biotech, Aidenbach, Germany
Dulbecco's Phosphate-buffered Saline (PBS)	PAN-Biotech, Aidenbach, Germany
Fetal bovine serum (FBS)	PAN-Biotech, Aidenbach, Germany
L-Glutamin	Biochrom GmbH, Berlin, Germany
Penicillin	Biochrom GmbH, Berlin, Germany
PolyJet	SignaGen Laboratories, Frederick, USA
Streptomycin	Biochrom GmbH, Berlin, Germany
Trypsin/EDTA (0,05%/0,02%)	Biochrom GmbH, Berlin, Germany

Table 5: Reagents used for cultivation of HEK23T cells

2.1.6. Buffers and solutions

Buffer/ solution	**Components**
Blotting buffer	10% methanol, 1x Sodium dodecyl sulfate (SDS)
DNA loading buffer (6x)	100 µl 6x loading buffer 150 µl Glycerol 1250 µl distilled H_2O
Sodium dodecyl sulfate (SDS) (10x)	250 mM Tris 1.9 M Glycine 1% Sodium dodecyl sulfate (SDS)
Radioimmunoprecipitation Assay (RIPA) buffer	25 mM Tris 137 mM Sodium chloride (NaCl) 0.1% Sodium dodecyl sulfate (SDS) 1% Glycerol 0.5% Sodium deoxycholate 1.0% Nonidet P-40 (NP-40) 2 mM Ethylenediaminetetraacetic acid (EDTA) pH 8.0
Tris-Acetate-EDTA Buffer (TAE) (20x)	0.8 M Tris 0.8 M Acetic acid 20 mM EDTA pH 8.0
Tris-buffered Saline with Tween 20 (TBST) (10x)	0.1 M Tris-HCl (Tris hydrochloride) 1.5 M Sodium chloride (NaCl) 0.5% Tween 20
Tris Hydrochloride (Tris-HCl)	1 M Tris Hydrochloride (Tris-HCl) pH 6.8
Tris Hydrochloride (Tris-HCl)	1,5 Tris Hydrochloride (Tris-HCl) pH 8.0

Table 6: Buffers and Solutions Used; µl: Microliter, M: Molar, mM: Millimolar.

2.1.7. Primary antibodies

Antibody	Detection of	Producer	Origin and clonality	Dilution
anti-Cas9 (7A9-3A3)	Cas9	sc-517386; Santa Cruz Biotechnology, Santa Cruz, USA	Mouse, monoclonal	1:500
Anti-HA	Hemagglutinin tags	MMS-101P; Covance, Münster Germany	Mouse, monoclonal	1:7500
Anti-Tubulin	Tubulin housekeeping gene	clone B5-1-2; Sigma-Aldrich, Taufkirchen, Germany	Mouse, monoclonal	1:7500
Anti-V5	V5 epitope tag of SIVagm Vif protein	MCA1360, ABDserotec, Düsseldorf, Germany	Mouse, monoclonal	1:4500

Table 7: Primary antibodies used for immunoblot

2.1.8. Secondary antibodies

Antibody	Description	Producer	Dilution
Anti- mouse-IgG-HRP	Sheep-derived antibody directed against mouse, conjugated to the reporter enzyme horseradish peroxidase	NA931V; GE Healthcare, Munich, Germany	1:10000
Donkey anti-Mouse IgG (H+L) - Alexa Fluor 488	Donkey-derived antibody directed against mouse, conjugated to the fluorescent dye Alexa Fluor 488	R37114, Thermo Fisher Scientific, Waltham, USA	3:1000

Table 8: Secondary antibodies used for immunoblot

2.1.9. Kits

Kit	Purpose	Producer
Amersham ECL Prime Western Blotting Detection	Signal detection of immunoblots	GE Healthcare Life Sciences, Munich, Germany
Pure Yield Plasmid MiniPrep	Small-scale plasmid DNA preparation from competent *E. coli* bacteria	Promega, Mannheim, Germany
Pure Yield Plasmid MaxiPrep	Large-scale plasmid DNA preparation from competent *E. coli* bacteria	Promega, Mannheim, Germany
QIAquick Gel Extraction Kit	Extraction of DNA fragments from agarose gel bands	QIAGEN, Hilden, Germany

Table 9: Kits used in this study

2.2. Methods

2.2.1. PCR-based creation of inserts for base editing plasmids

Plasmids are small, circular double-strand DNAs (dsDNA) that are commonly used in molecular biology to carry foreign genes or sequences for protein expression within a host cell. The plasmids A3Ai-Cas9n-UGI-NLS (Addgene #109425; Addgene, Cambridge, USA), MLM3636 sgRNA construct (Addgene #43860), and plenti-CMV-mCherry-T2A-GFP (Addgene #109427) were obtained from Addgene.

Polymerase Chain Reaction (PCR) is a widely used technique that enables the amplification of specific segments of DNA. It involves the use of short DNA primers that flank the target region, along with a heat-stable DNA polymerase enzyme to replicate the DNA. PCR allows for the rapid amplification of DNA sequences, which can then be used for cloning. The PCR reaction is given in table 10.

Reagent	**Volume**
DNA template	X µl (10-100 ng)
Forward primer (10 µM)	2,5 µl
Reverse primer (10 µM)	2,5 µl
dTNPs (10 mM)	1 µl
Q5 DNA-polymerase	0,5 µl
Q5 DNA-polymerase buffer	10 µl
Nuclease-free water	To total volume 50 µl

Table 10: Reagents and volumes required for PCR reactions; µM: micromolar, µl: microliter, mM: milimolar

For the amplification the T3 Thermocycler (Biometra Biomedizinische Analytik, Jena, Germany) was used. Unless stated otherwise the PCR program was the same for all reactions and is given below.

Initial Denaturation: 98°C for 2 minutes

Denaturation: 98°C for 30 seconds

Annealing: X°C for 30 seconds (X= optimal annealing temperature depending on primer sequence and GC content)

Extension: 72°C for Y seconds (Y = ~1 kb per 15 seconds for Q5 high-fidelity polymerases, i.e. the time required varied according to the size of the desired insert)

Repeat Steps 2–4: 25–35 cycles

Final Extension: 72°C for 2 minutes

Hold: 4°C indefinitely

The primers and their sequences are all listed below in table 11.

Primer name	Sequence (5’ to 3’ direction)	Purpose
huA3C FP	AGATCCGCGGCCGCGCCGCCACCATGAA TCCACAGATCAGAAACCCGATGA	huA3C and variants cloning
huA3C RP	TGAGGTCCCGGGAGTCTCGCTGCCGCTC TGGAGACTCTCCCGTAGCCTTCTT	huA3C and variants cloning
S188I A3C RP	TGAGGTCCCGGGAGTCTCGCTGCCGCTC TGGAGAATCTCCCGTAGCCTTCTT	huA3C S188I cloning
Ext. common αHe1 FP	GTGATGCGGTTTTGGCAGTA	A3C αHelix1 and Loop1 variants cloning
Ext. common αHe1 RP	GGTCCCGGGAGTCTCGCTGC	A3C αHelix1 and Loop1 variants cloning
Int. CH2-αHe1 FP	GATCCACACATATTCTACTTCCACTTTAAA AACCTATGGGAAGC	A3C αHelix1 variants cloning
Int. CH2- αHe1 RP	GTGGAAGTAGAATATGTGTGGATCCATT GCCTTCATCGGGTTTC	A3C αHelix1 variants cloning
Int. hu- αHe1 FP	TATCCAGGCACATTCTACTTCCAATTTAA AAACCTACGGAAAGC	A3C αHelix1 variants cloning
Int. hu- αHe1 RP	TTGGAAGTAGAATGTGCCTGGATACATT GCCTTCATCGGGTTTC	A3C αHelix1 variants cloning
Int.huA3C.He1 CH2.RK FP	TAAAAACCTACGGAAAGCCAACGATC	A3C Loop1 variants cloning
Int. huA3C.He1 CH2.RK RP	GATCGTTGGCTTTCCGTAGGTTTTTA	A3C Loop1 variants cloning
Int.huA3C.He1 CH2.YG FP	ATGGGAAGCCTATGGTCGGAACGAAA	A3C Loop1 variants cloning
Int. huA3C.He1 CH2.YG RP	TTTCGTTCCGACCATAGGCTTCCCAT	A3C Loop1 variants cloning

GFP L138S FP	GGACGGCAACATTTCAGGGCACAAGCTGGA	GFP L138S mutation
GFP L138S RP	TCCAGCTTGTGCCCTGAAATGTTGCCGTCC	GFP L138S mutation
GFP L202S FP	CGACAACCACTATTCAAGTACCCAGTCGGCCCTGA	GFP L202S mutation
GFP L202S RP	TCAGGGCCGACTGGGTACTTGAATAGTGGTTGTCG	GFP L202S mutation
GFP L138S gRNA FP	ACACCCAACATTTCAGGGCACAAGCG	GFP L138S gRNA cloning
GFP L138S gRNA RP	AAAACGCTTGTGCCCTGAAATGTTGG	GFP L138S gRNA cloning
GFP L202S gRNA FP	ACACCCCACTATTCAAGTACCCAGTG	GFP L202S gRNA cloning
GFP L202S gRNA RP	AAAACACTGGGTACTTGAATAGTGGG	GFP L202S gRNA cloning
NS gRNA FP	ACACCGCACTACCAGAGCTAACTCAG	NS gRNA cloning
NS gRNA RP	AAAACTGAGTTAGCTCTGGTAGTGCG	NS gRNA cloning
T2A GFP FP	TGCAGATCTCGAGTACAAGGAGGGCAGAGG	T2A GFP cloning
T2A GFP RP	TCTTAAAGGTACCTTATTTATATAATTCATCCAT	T2A GFP cloning

Table 11: Primers used for inserts

2.2.1.1. Site directed mutagenesis for eGFP reporter plasmid

eGFP L138S and eGFP L202S reporter constructs were created using site-directed mutagenesis. This technique uses primers designed with specific mutations in their sequence, which allows for the incorporation of precise point mutations into the amplified DNA. Once the inserts were obtained, the DNA fragment and the reporter construct plasmids were subjected to restriction digestion (see below).

2.2.1.2. Creating sgRNAs with the MLM3636 vector

The MLM3636 plasmid is a vector specifically designed to serve as guide RNA. All sgRNA coding sequences were cloned according to the manufacturers protocol. The ssDNA oligonucleotides encoding the L202S and L138S sgRNAs and the non-specific RNA (nsRNA) were suspended to 100 µM in 0.1xTE buffer and then diluted 1:10 to 10 µM in H_2O. In order to generate the phosphorylated oligo duplex, 1 µL top strand oligonucleotide (10 µM), 1 µL bottom strand oligonucleotide (10 µM), 2 µL 10x T4 DNA Ligase buffer (New England Biolabs, Ipswich, USA), 0.5 µL T4 PNK (New England Biolabs, Ipswich, USA), and 15.5 µL H_2O were mixed. The incubation was carried out using the GeneTouch thermocycler (Bioer Technology, Hangzhou, China) under the following conditions: after 60 minutes at 95°C, the reaction was heated up to 95°C for 5 minutes. The temperature was then ramped down to 10°C at -5°C per minute.

2.2.1.3. Creating A3C variant base editor plasmids from pcDNA templates

Mutants of huA3C, including A3C.WE-RK, A3C S188I, A3C CH2, and smmA3C-like in pCDNA (Thermo Fisher Scientific, Waltham, USA) vectors served as templates to generate the A3C editing constructs. The pcDNA3.1 (+/-) is a common expression plasmid, the A3C variants were previously created by former members of our group. Since there was already a template with the desired sequence available, primers flanking the A3C sequence were used to create the insert in one PCR reaction (see above).

2.2.1.4. Overlapping PCR for A3C αHelix1 and loop1 variant base editor plasmids

A3C αHelix1 and loop1 variant inserts were generated through overlapping PCR. Overlapping PCR requires a total number of three PCR reactions. A first round of PCR (two independent reactions) creates two DNA fragments with overlapping sequences. Thereafter, the two fragments are combined in a second round of PCR, creating the desired insert.

2.2.2. Agarose Gel Electrophoresis of PCR products

To verify whether the DNA amplified through PCR corresponded to the desired fragment length, agarose gel electrophoresis was performed. Agarose gel electrophoresis is used to separate DNA fragments based on their size. DNA is negatively charged and migrates from the cathode to the anode in an electric field, with the gel matrix acting as a filter. Nucleic acid strands move through the agarose gel at different speeds depending on their size, allowing for separation.

As a size standard, the GeneRuler 1 kb Plus DNA ladder (Thermo Fisher Scientific, Waltham, USA) was used. In this study, 1% agarose gels were prepared by dissolving agarose in Tris-acetate-EDTA (TAE) buffer in a microwave. After cooling down, ethidium bromide solution was added, which intercalates with DNA and fluoresces under ultraviolet (UV) light.

After polymerization, the gel was placed in a gel chamber filled with TAE buffer and loaded with nucleic acid samples mixed with DNA loading buffer, along with the DNA ladder. Electrophoresis was carried out at 110 V for 20–30 minutes. The DNA bands were visualized using a UV transilluminator.

If the visualized DNA band matched the desired length, the piece of gel containing the band was excised and purified using the QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany). 200 µl of QG buffer were added to the slice of gel and incubated at 50°C until the gel had completely dissolved. All centrifugation steps were carried out at $13x10^3$ rpm for 1 minute. The mixture is then transferred to a QIAquick spin column placed into a 2 ml collection tube and then centrifugated. After discarding the flowthrough, 710µl of PR buffer were added twice. In between adding the buffer, the column was centrifugated and the flowthrough removed. The DNA was eluted in 20-50 µl of elution buffer and gathered in a clean tube through centrifugation.

2.2.3. Restriction digestion of PCR products and vector backbones

PCR products and circular vector backbones were digested with restriction enzymes to create compatible ends for ligation. Restriction sites are specific DNA sequences recognized by restriction enzymes. The restriction enzymes are able to cut the DNA at

these precise locations. This allows for the insertion or removal of DNA fragments, facilitating cloning and genetic modifications. Digestion linearizes the circular plasmid, preventing self-ligation and increasing cloning efficiency. Furthermore, it allows for the removal of unwanted DNA fragments within the vector and the precise insertion of the PCR product at the multiple cloning site. Using two different restriction enzymes helps ensure the correct orientation of insert. Unless stated otherwise, 2 µl of each restriction enzyme, 3 µl of the respective buffer were mixed with individually calculated amounts of DNA. Nuclease-free water was added to all reactions to adjust to a final volume of 30 µl. The individual reactions for restriction digestions are given below.

2.2.3.1. Restriction digestion for the eGFP reporter plasmid

Once the PCR products (see above) were obtained, the DNA fragment and the reporter construct plasmids were subjected to restriction digestion. XhoI and KpnI were used as restriction enzymes. As both enzymes require a specific, non-compatible buffer for an optimal reaction, two individual digestion steps were performed. For the insert digestion 18 µl of PCR product was mixed with 2 µl XhoI enzyme, 3 µl buffer R and 7 µl nuclease-free water to a total volume of 30 µl. The vector digestion was carried in the following reaction: 2000 ng of reporter plasmid (i.e., 2,8 µl), 2 µl XhoI enzyme, 2 µl buffer R, and 13 µl nuclease-free water were mixed to obtain a final volume of 20 µl. Both reactions were incubated at 37°C for 1 hour.

Afterwards the digested products were purified using the Qiagen QIAquick Gel purification kit and eluted with 20 µl elution buffer. This step was followed by the second digestion using the KpnI restriction enzyme. As both insert and product were eluted in 20 µl, the subsequent restriction digestion was the same for both products. 20 µl of eluted insert or vector was mixed with 2 µl of KpnI, 3 µl Kpn buffer, and 5 µl of nuclease-free water for a final volume of 30 µl. After an incubation period of 1 hour at 37°C, the insert was purified once more using the Qiagen QIAquick Gel Purification kit. The vector, however, had been cut into two pieces, one with a length of roughly 100 bp and a second one with a length of 800 bp, which is the fragment of interest. Therefore, an agarose gel as described in 2.2.2. was run in order to separate the two fragments. After running and

cutting the gel accordingly, the band was purified using the QIAquick Gel Extraction Kit. Afterwards the DNA concentration was measured he NanoDrop 200 1.6 software (Thermo Fisher Scientific, Waltham, USA). This method involves a photometric measurement of light absorption at a wavelength of 260 nm, performed using the NanoDrop NP-1000 spectrophotometer (Peqlab, Erlangen, Germany).

2.2.3.2. Restriction digestion for the sgRNA plasmids

The vector backbone was digested using the BsmBI restriction enzyme (Thermo Fisher Scientific, Waltham, USA). For the digestion, 2 µg plasmid DNA were mixed with 2 µl BsmBI enzyme, 4 µl Tango buffer (Thermo Fisher Scientific), and nuclease-free water to sum to a final reaction volume of 40 µl. After an incubation period of 4 hours at 55°C, the cut vector backbone was isolated using the Qiagen QIAquick Gel Purification Kit and the DNA concentration was measured.

2.2.3.3. Restriction digestion for all base editor constructs

The plasmid A3Ai-Cas9n-UGI-NLS served as vector backbone. To create the A3C-based editing constructs, the A3Ai sequence had to be removed. This was achieved using Cfr9I (XmaI) and NotI restriction sites within the plasmid. As Cfr9I (Thermo Fisher Scientific) requires a Cfr9I-specific buffer, NotI (Thermo Fisher Scientific) volume was doubled to ensure efficient digestion in a single reaction. 200 ng of plasmid were digested with 1 µl Cfr9I enzyme, 2 µl NotI, 2 µl Crf9I-specific buffer, and nuclease-free water up to a final volume of 20 µl. The reaction was incubated at 65°C for two hours. Following digestion, agarose gel electrophoresis was performed, and the empty vector backbone was purified.

In parallel, the purified insert was also double-digested under the same conditions.

2.2.4. Insert vector ligation

After restriction digestion, the insert and linearized vector were ligated. The required amounts of insert and vector were determined based on insert length, vector length, and vector mass, using the NEBioCalculator (New England Biolabs, Ipswich, USA). All

ligation reactions were performed following the same protocol and are therefore not stated individually. Ligations were performed at insert-to-vector ratios of 2:1. The reaction mixture contained the calculated amounts of insert and vector, 1 µl T4 DNA ligase (New England Biolabs, Ipswich, USA)., and 1 µl T4 DNA Ligase Buffer (New England Biolabs, Ipswich, USA), adjusted to a final volume of 10 µl with nuclease-free water. The mixture was incubated at room temperature for 1–2 hours, after which the ligation product was directly transformed into competent *Escherichia coli* (E. coli) bacteria.

2.2.5. Amplification of plasmid DNA

2.2.5.1. Transformation in chemically competent bacteria

For plasmid amplification, strains of *E. coli* bacteria were transformed with the respective plasmids. The term transformation refers to the non-viral introduction of free DNA into competent bacteria. Bacterial strains Top10 and Stabl2 (Life Technologies, Darmstadt, Germany) were used for plasmid propagation. Prior to use, the bacteria had been made chemically competent following a modified version of Hanahan's method (Hanahan, 1983).

The bacteria were cultured in LB medium. The components listed in table 11, except for ampicillin, were mixed. After adjusting the pH to 7 using sodium hydroxide, the medium was autoclaved for 20 minutes. Once the medium reached a temperature below 50°C, the selection antibiotic ampicillin was added.

Component	**Solid medium**	**Liquid medium**
ddH_2O	1 l	1 l
Caseinpepton	10 g	10 g
Yeast extract	5 g	5 g
Sodium chloride (NaCl)	10 g	10 g
Agar	15 g	-
Ampicillin	50 mg	50 mg

Table 11: Components of the LB medium; g: gram, l: liter, mg: milligram

For the transformation, the *E. coli* stored at -80°C were thawed on ice and then mixed with 20–50 ng of plasmid per 100 µl of bacterial suspension. After an incubation period

30 minutes on ice, the suspension was subjected to a heat shock at 42°C for 45-70 seconds. The heat shock was followed by another 5–10-minute incubation on ice. Thereafter 800 µl of LB medium was added, and the suspension was incubated on a shaker at 550 rpm for 60–90 minutes and 30°C for the Stbl2 and 37°C for the Top10 strain. Post incubation the bacteria were centrifuged at 3,3 rpm for 5 minutes and then spread onto agar plates containing ampicillin.

After 24–48 hours of incubation at 30°C or 37°C (depending on the strain as described above), a single bacterial colony was picked from the plate and transferred into LB liquid medium. Depending on the purpose of plasmid propagation, the colony was either transferred into 4 ml or 250 ml of LB medium supplied with Ampicillin. The culture was then incubated with agitation at 150-180 rpm for an additional 24–48 hours at 30°C or 37°C respectively.

2.2.5.2. DNA preparation from competent *E. coli*

For DNA preparation from competent bacteria, either the PureYield™ Plasmid Miniprep System (Promega, Mannheim, Germany) or the PureYield™ Plasmid Maxiprep Kit (Promega, Mannheim, Germany) was used according to the manufacturer's instructions.

2.2.5.3. Small-scale DNA preparation

Small-scale DNA preparation usually was applied for newly created plasmids following insert vector ligation. When subjected to the Miniprep kit, 1.5 ml of bacterial suspension was centrifugated at $10x10^3$ rpm for 1 minute. After centrifugation, the supernatant was removed, more bacterial suspension added and the centrifugation process repeated. The repetition allows for higher plasmid concentrations and purity. Once the supernatant was removed for the second time, 100 µl of cell lysis buffer was added and the contents of the tube mixed by inverting. After adding 350µl of cold neutralization solution and mixing thoroughly, the tubes were centrifugated at full speed for 3 minutes. Thereafter, the supernatant was carefully transferred to a filtering column, the column placed into a collection tube and then centrifugated at full speed for 15 seconds. After discarding the flowthrough, 200 µl of endotoxin removal solution was added and passed

through the column by centrifugating at maximum speed for 15 seconds. 400 µl of a column wash solution were added and followed by another centrifugation at maximum speed for 30 seconds. In order to finally elute the DNA, 30 µl of nuclease-free water were added to the column and centrifugated at full speed for 15 seconds. DNA concentration was determined using the NanoDrop 200 1.6 software and NanoDrop NP-1000 spectrophotometer. Finally, the eluted DNA was stored at -20°C. Newly created and extracted plasmids were sent for sanger sequencing to confirm a successful PCR insert production and cloning process. Sanger sequencing is a method used to determine the exact order of bases in a DNA strand.

2.2.5.4. Large-scale DNA preparation

Large-scale DNA preparation was used on plasmids with confirmed sequences. For the Maxiprep kit, the bacterial suspension was centrifuged at 5000 × g for 10 minutes at 22°C and the supernatant was discarded. The remaining cell pellet was resuspended in 12 ml of resuspension solution by pipetting up and down and briefly vortexing. Afterwards, 15 ml of cell lysis solution was added and the flaks were gently shaken until the content developed a clumpy consistency. Following an incubation period of 3 minutes at RT, 15 ml of neutralization solution was added and the flasks were inverted 15-20 times. Once the suspension became more homogeneous, it was centrifugated at 7000 × g for 30 minutes. The supernatant was then passed through a filtering column under vacuum pressure onto a column with a DNA-binding membrane. After passing the liquid, the membrane was washed with 5 ml of endotoxin removal solution and then with 20 ml of ethanol-containing wash solution. The purified DNA was eluted from the membrane using nuclease-free water and collected in microcentrifuge tubes. Afterwards the DNA concentration was measured as described above and the plasmid DNA stored at -20°C.

2.2.6. Cell culture methods

Work in cell culture was exclusively conducted using a laminar flow cabinet within a security level S2 laboratory. Adherent cells were cultivated in Dulbecco's High-Glucose

Modified Eagle's Medium (DMEM, PAN-Biotech, Aidenbach, Germany) completed with serum and antibiotics (see table 12.) at 37°C and 5% CO2.

Supplement	**Required amount**
Fetal bovine serum (FBS)	10%
L-glutamine	2 mM
Penicillin	100 U/ml
Streptomycin	50 mg/ml

Table 12: Supplements for complete DMEM medium; mg: milligram, mM: millimolar, ml: milliliter, U: unit

HEK293T cells were split every two to seven days. Splitting the cells first included removing the medium and washing the cells with 7 ml of Dulbecco's Phosphate-buffered Saline (PBS, PAN-Biotech, Aidenbach, Germany). Subsequently they were treated with 2.5 ml of Trypsin/EDTA (0.05%/0.02%, Biochrom GmbH, Berlin, Germany) to detach the cells from the bottom of the flask. After 5 minutes of incubation at room temperature 7.5 ml of complete DMEM were added and the cells resuspended using a pipettor. According to demand, cells were split at different ratios ranging between 1:5 and 1:20 and returned to their flasks. When used for experiments, the cells were first counted and then seeded in plates. To conserve the cells, they were frozen and stored in a liquid nitrogen tank.

2.2.6.1. Transfection for episomal base editing

A transfection by definition is the introduction of foreign genetic material to eukaryotic cells. For the base editing experiments different plasmids were transiently transfected in both WT HEK293T and transduced HEK293T (see below). Cells were seeded in 6-well plates, per well $6 \bullet 10^5$; exceptions are explained later on. 24 hours after seeding, a maximum of 1500 ng plasmid and 4 µl PolyJet reagent (SignaGen Laboratories, Frederick, USA) were mixed with 200 µl of blank DMEM. After incubating the mixture for 15 minutes at room temperature it was carefully added to the cells. PolyJet is a transfection reagent designed to introduce nucleic acids, such as plasmid DNA or into mammalian cells. PolyJet acts through forming complexes with negatively charged nucleic acids, creating a positively charged particle that can interact with the cell membrane. This

interaction facilitates the uptake of plasmid into the cell through endocytosis. Once inside the cell, the nucleic acids reach the nucleus and transcribed, ultimately allowing for the expression of the desired protein. For episomal base editing three plasmids were co-transfected: 200 ng sgRNA, 400 ng reporter and 600 ng base editor. The plasmids used for transfection are listed above Following 48 hours of incubation the cells expressed the desired proteins and were used for further base experiments, such as flowcytometry or immunoblot.

2.2.6.2. Production of viral particles

For the transduction of HEK293T cells for chromosomal BE, viral particles were required. In order to obtain such particles, the lentiviral L138S and L202S reporter plasmids were co-transfected with a VSV-G and Rev, Gag-Pol expression plasmids. The reporter plasmids carry long terminal repeats (LTR), which are essential for the integration of the desired genes in the cell genome. The other two plasmids are required for the creation of viral particles. The VSV-G plasmid expresses the membrane-associated glycoprotein of the VSV. Gag encoded the viral particles' structural proteins and Pol the viral enzymes that ultimately allows for the integration of the genes of the reporter plasmids in the cell genome. Finally, REV is required for the exportation from the nucleus the messenger RNA (mRNA).

48 hours after transfection, the supernatant was collected and centrifuged at 4 °C for 5 minutes at a speed of 5500 rpm to remove cellular components. The collected particles were then stored at -80°C or directly used for transduction. For the viruses and the plasmids used, see the table below.

Virus	**Plasmids**	**Genes**	**Transfected amount**
L202S reporter	L202S eGFP reporter plasmid	eGFP, mCherry	500 ng
	psPAX2	*gag*, *pol*, *rev*, *tat* RRE	900 ng
	pMD.G	VSV-G	200 ng

L138S reporter	L138S eGFP reporter plasmid	eGFP, mCherry	500 ng
	psPAX2	*gag*, *pol*, *rev*, *tat*, RRE	900 ng
	pMD.G	VSV-G	200 ng
Table 13: Plasmids used for the production of viral particles; ng: nanogram			

2.2.6.3. Transduction for reporter cells and transfection for chromosomal BE

To obtain reporter cells, i.e., cells that constitutively express the GFP reporter, HEK293T cells were transduced using 500 µl of either of the reporter viral particles. 48 hours after transduction, the mCherry positive cells (thus successfully transduced) were retrieved by fluorescence-activated cell sorting (FACS) and cultured with complete DMEM. Since the transduced cells were already selected through FACS, no selection antibiotic was required. Following selection, the cells were seeded and transfected similarly to episomal base editing experiments. The amount of transfected sgRNA remained the same, merely the transfected amount of base editor was increased to 800 ng. A schematic depiction of the experimental setup is given in fig. 5.

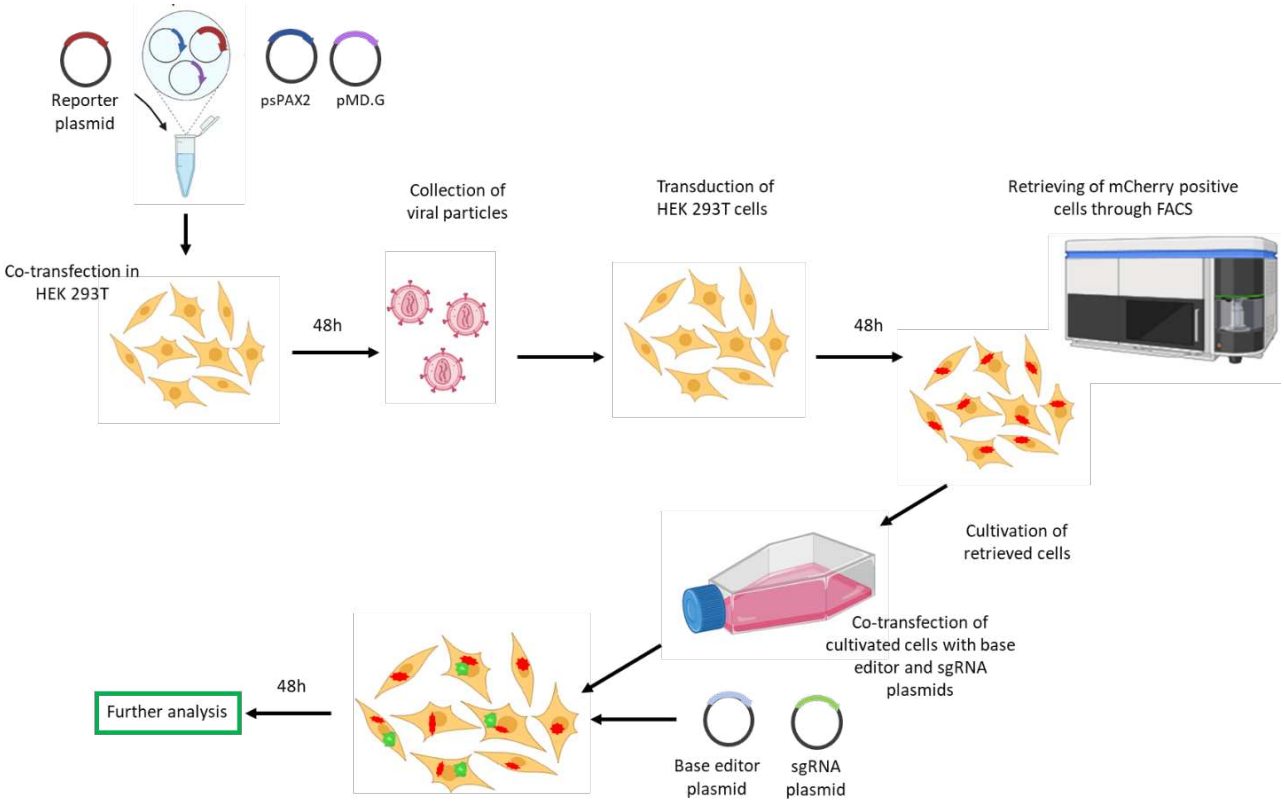

Figure 5. Graphical illustration of the production of reporter cells. The co-transfection of the reporter plasmids with psPAX2 and pMD.G in HEK 293T cells allows for the creation of viral particles. After 48 hours the viral particles are collected and used for the transduction of HEK 293T cells. Following an incubation period of 48 hours, successfully transduced cells are retrieved through FACS. These mCherry positive cells were then taken into cell culture and ultimately co-transfected with plasmids for the base editors and sgRNA. After 48 hours the cells were then subjected to further experiments, such as flowcytometry or immunoblot analysis.

2.2.7. Flow cytometry and FACS

Flow cytometry is a technique used to analyze and sort cells based on their physical and chemical properties. The process involves suspending cells in a fluid and passing them through a laser beam, one cell at a time. Usually, cells are labeled with fluorescently tagged antibodies or dyes, which bind to specific cell components or markers. However, since the transfected or transduced HEK293T cells already expressed mCherry and potentially eGFP, no additional fluorophores were needed.

Flow cytometers are equipped with multiple lasers, each emitting light at specific wavelengths. These lasers excite the fluorophores both on the cell surface or within the cell. For example, a blue laser (around 488 nm) may excite green fluorescent proteins (i.e., eGFP), while a yellow-green (typically 561 nm) may be used to detect red fluorescence, such as of mCherry. As the cells pass through the laser beam, the light they scatter and the emitted fluorescence are detected by sensors. Forward scatter (FSC) measures the light scattered in the forward direction and correlates with the cell size. Side scatter (SSC) measures light scattered at a 90-degree angle, providing information on the cell's internal complexity or granularity. This allows for the typification of cells. The fluorescence emitted by the tags is captured by photomultiplier tubes or other detectors, which are tuned to specific wavelengths based on the emission spectra of fluorophores used. The data collected from light scatter and fluorescence is then converted into digital signals for analysis. Data is usually displayed as histograms or dot plots and enables the immediate assessment cell characteristics, such as the presence of specific markers, size, or complexity.

Flowcytometry enabled the real-time quantification of BE frequencies within the target cells based on the expression of mCherry and eGFP. Transfected HEK293T cells were harvested and resuspended in FACS buffer (2 mM EDTA, 2% FBS in 1x PBS sterile filtered). Cells were analyzed at a 488 nm excitation wavelength using a FACSAria III cell sorter (BD Biosciences, Heidelberg, Germany).

Flowcytometry data was then analyzed using the FlowJo version 10.4 software. The same gating strategy was applied across all flow cytometry and FACS experiments. First, the 293T cell population was determined after the exclusion of debris based on the FSC-Area (FSC-A) and SSC-Area (SSC-A). This population was then subjected to single cell gating to ensure downstream analysis of individual cells. This was achieved by gating the diagonal population in the FSC-Height (FSC-H) versus FSC-A plot, effectively removing doublets. eGFP and mCherry fluorescence were subsequently measured within the single-cell population to assess BE efficiency. Based on the fluorophores' respective emission wavelengths, detection was performed using the FITC and PE-Texas Red channels.

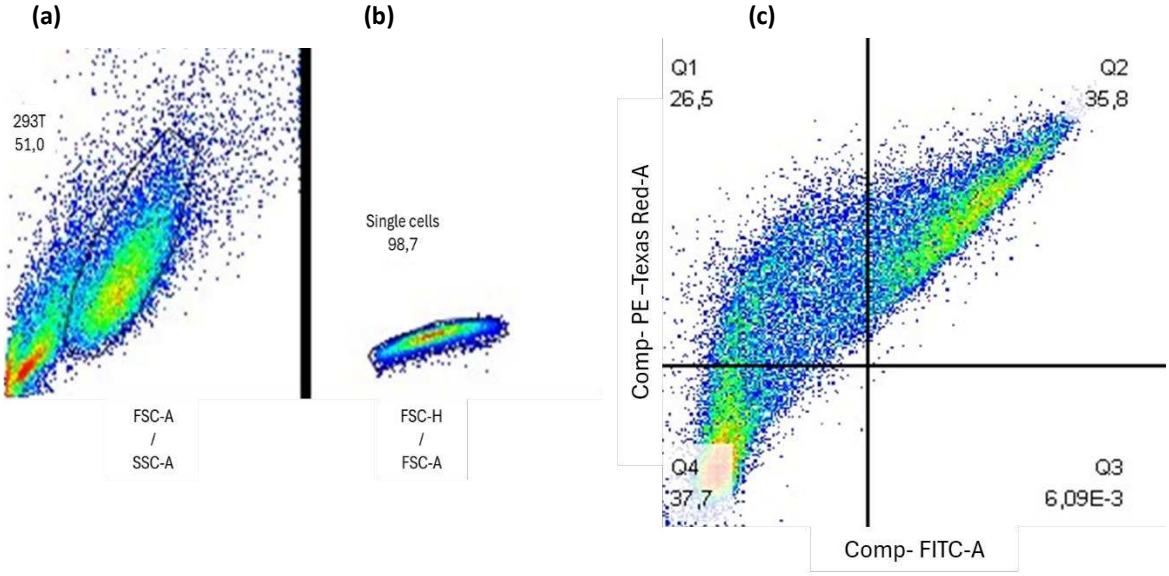

Figure 6. Representative image of gating strategy. (a) 293T cell population were first gated by FSC-A versus SSC-H to exclude debris. 51% of the counted events were included in this population. (b) Single cells (98,7% of the 293T cells) were then selected using FSC-H versus FSC-A to remove doublets. (c) eGFP and mCherry fluorescence were measured within this single-cell population using the FITC and PE-Texas Red channels, respectively. Double positive cells (shown in Q2) represent the effectively edited cells.

For chromosomal BE experiments a pool of cells stably expressing the reporter plasmids was required. Therefore, transduced HEK293T cells were selected by the FACSAria III cell sorter. FACS is a specialized type of flow cytometry that allows for both the analysis and physical sorting of cells based on their physical properties and fluorescence characteristics. Cells are analyzed as described above and those meeting selection criteria are encapsulated into small, electrically droplets. These droplets are then deflected into separate collection tubes by an electric field, based on their charge. This enables the effective sorting of the cells into different populations. Effectively transduced HEK293T cells express mCherry as they carry the reporter plasmids. Similar to the flowcytometry on transfected HEK293T cells, transduced cells were collected and resuspended in FACS buffer. Cells were then sorted based on whether they were positive for mCherry or not. To prevent cell death, cells were kept on ice prior to sorting. As the sorting process can be highly time extensive and prolonged sorting time decreases the chance of cell survival, initially small cell populations were created. Populations ranging between 30.000-100.000 cells were taken into cell culture to be grown and cultivated. Upon sufficient cell growth they could be used for further BE experiments.

2.2.8. Cell imaging

Episomal and chromosomal BE experiments were conducted using an mCherry-eGFP reporter assay. While successful transfection or transduction led to the expression of mCherry (red fluorescence), BE activity was reflected in the expression enhanced green fluorescence (eGFP). The emitted signals allowed the real-time assessment of BE experiments.

2.2.8.1. Fluorescence microscopy

In cell culture, the microscope AE29 (Motic, Barcelona, Spain) was used to observe the expression of the fluorescent proteins. Transfected and transduced cells were checked on routinely every 12-24 hours. Fluorescence microscopy images were acquired using the XY cell observer and analyzed ZEN lite (black edition) software (Carl Zeiss).

2.2.8.2. Confocal microscopy

For confocal microscopy, 1x 10^5 HEK293T cells were seeded on polyethene coverslips (Thermo Fisher Scientific, Vilnius, Lithuania) and transfected with 500 ng of base editing construct the following day. Confocal microscopy is a technique that uses a focused laser beam to scan a sample and detect emitted light from a specific focal plane, allowing for high-resolution imaging. It enables studying fluorescently labeled molecules whilst offering clearer, more detailed views of cellular structures.

48 hours post-transfection, cells were first fixated with 500 µl of 4% paraformaldehyde in PBS for 10 minutes and then washed twice with PBS. All washing steps were carried out using PBS. Cells were then permeabilized with 500 µl 0.1% Triton X-100 in PBS for 10 minutes at 27°C and washed three times. Thereafter, the cells were blocked with 500 µl of 10% FBS in PBS (blocking buffer) for 1 hour and once more washed. Subsequently, the primary antibody, anti-Cas9 (see above; dilution 1:500 in blocking solution), was added for 1 hour at 27°C. After removing the primary antibody and washing three times, the secondary antibody, donkey anti-mouse Alexa Fluor 488 (Covance, Münster, Germany), diluted in blocking solution (1:300) was added. The cells were incubated for 1 hour at 27°C and then washed three times. Finally, 4',6-diamidino-2-phenylindole (DAPI) (Thermo Fisher Scientific) was used to stain the cell nuclei. The stain was diluted 1:1000 in blocking solution and applied for 2 minutes at 27°C. After removing the DAPI stain and washing three times, the coverslips were mounted onto slides.

Finally, images were obtained by a 60× objective using a Zeiss LSM 510 Meta laser scanning confocal microscope (Carl Zeiss, Cologne, Germany). Captured images were subsequently analyzed by ZEN lite (black edition) software (Carl Zeiss).

2.2.9. Protein expression analysis and detection

2.2.9.1. Preparation of protein lysates from eukaryotic cells

To analyze protein expression in cells, immunoblots were performed. Immunoblots first require the preparation of protein lysates. Usually, 1.2 x 10^6 cells per well were seeded in 6-well plates and 48 hours later the medium was removed, the cells washed with and

collected in 2–3 days later, the medium was removed, and the cells were collected in PBS. Cells were then pelleted through centrifugation at $5x10^3$ rpm for 5 minutes.

Next, cells were lysed into radioimmunoprecipitation assay (RIPA) buffer (25 mM Tris-HCl [pH7.6], 150 mM NaCl, 1% NP-40, 1% sodium deoxycholate, 0.1% sodium dodecyl sulfate [SDS] mixed with protease inhibitor cocktail set III [Calbiochem, Darmstadt, Germany]). In order to lysate the cells properly with the buffer, the mixture was pipetted up and down and then incubated on ice for 5 minutes. Thereafter, cells were centrifugated at maximum speed for 20 minutes at 4°C and the supernatant was transferred into a clean tube.

2.2.9.2. Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

SDS-PAGE allows for the separation of proteins based on their molecular weight, which in turn leads to conclusions about the protein expression in cells. The lysates prepared as describe above were treated with the SDS-containing loading buffer RotiLoad (Carl Roth, Karlsruhe, Germany), which contains β-mercaptoethanol to break covalent bonds in the secondary structure of proteins.

SDS is anionic and ensures that all proteins are negatively charged enabling the molecular weight-dependent separation of proteins during electrophoresis. For every 60 µl of lysate, 20 µl of RotiLoad was added. The protein samples were then denatured at 95°C for 5 minutes.

The resolution of protein separation depends on the pore size of the running gel, which is determined by the acrylamide concentration. In this study, stacking gels with 5% acrylamide and separating gels with 8% acrylamide were used (see tables 14 and 15). The purpose of the stacking gel is to concentrate the proteins before they pass through the separating gel based on their molecular weight.

Component	**Required amount**
30% Acrylamid 0.8% Bisacrylamid	840 µl

1 M Tris base, pH 6.8	630 µl
10% SDS	50 µl
ddH_2O	3.5 ml
20% APS	25 µl
TEMED	5 µl

Table 14: Stacking gel- 5 % acrylamide (2 gels); µl: microliter, ml: milliliter

Component	**Required amount**
30% Acrylamid 0.8% Bisacrylamid	4 ml
1.5 M Tris base, pH 8.8	3.75 ml
10% SDS	150 µl
ddH_2O	7.1 ml
20% APS	75 µl
TEMED	15 µl

Table 15: Running gel – 8% acrylamide (2 gels); µl: microliter, ml: milliliter

Gels were prepared as described as in the given tables, tetramethylethylenediamine (TEMED) and ammonium persulfate (APS) enable polymerization. First, the running gel was pipetted between two cleaned glass plates fixed in a clamp holder. The gel surface was covered with a thin layer of isopropanol ensuring an even polymerization. Once solidified, the isopropanol was removed and the stacking gel was added. A comb was inserted into the stacking gel to create wells for sample to be loaded.

After loading the protein samples and the PageRuler Prestained Protein Ladder (Thermo Fisher Scientific, Waltham, USA) as a molecular weight marker, a voltage of 85 V was applied until the proteins were concentrated at the bottom of the stacking gel. The voltage was then increased to 110 V and maintained until the proteins were sufficiently separated.

Following electrophoresis, the gel was removed from the glass plates and the stacking gel was discarded.

2.2.9.3. Immunoblot

The proteins separated were then transferred onto a polyvinylidene fluoride (PVDF) membrane (Merck Millipore, Darmstadt, Germany). The subsequent immunoblot technique allows for the protein detection with antibodies. First, the membrane was activated in methanol. In this study, a semi-dry transfer method was employed, where thick filter paper was soaked in blotting buffer and placed on the anode of the transfer unit. The activated PVDF membrane was washed in blotting buffer and then placed on top of the filter paper. The cut gel was placed on the membrane, followed by another layer of thick filter paper soaked in blotting buffer. Air bubbles were removed at each step. The transfer unit was closed with the cathode-bearing lid and a voltage of 25 V was applied for a minimum of 45 minutes. This enabled the negatively charged proteins to transfer from the gel to the membrane.

2.2.9.4. Immunodetection of Proteins

Proteins bound to the PVDF membrane were detected and visualized using specific antibodies. After blotting, the membrane was incubated for 1 hour in Tris-buffered saline with Tween 20 (TBST) containing 5% milk powder. This blocking step prevents the antibody from non-specific binding. The membrane was then incubated with a primary antibody overnight at 4°C. The membrane was incubated either with agitation or rotation to ensure an even distribution of the antibody. The antibodies used are listed above in table 7. The primary antibodies were diluted in TBST with 5% milk powder at the specified concentrations. The following day, the membrane was washed three times with TBST for 5-10 minutes each. After washing, the membrane was incubated for 1 hour with the secondary antibody (table 8), which binds to the Fc region of the primary antibody. The secondary antibody was conjugated to horseradish peroxidase, a detection enzyme. Upon adding the Enhanced chemiluminescent luminol-based (ECL) substrate (GE Healthcare Life Sciences, Freiburg, Germany) the reaction of the contained luminols was catalyzed and a light signal became detectable. In this study, the chemiluminescence

was captured using photographic films and made visible with the aid of a film development apparatus (AGFA Healthcare, Mortsel, Belgium).

2.2.9.5. Immunoprecipitation for A3-Cas9

To investigate APOBEC-Cas9 protein interaction, a co-immunoprecipitation (Co-IP) assay was performed. Co-IP is a technique used to study protein-protein interactions by isolating a target protein along with its binding partners using antibody-tagged beads. After immunoprecipitation, the co-precipitated proteins are analyzed through immunoblot.

For this purpose, HEK293T cells were either singly or co-transfected with A3 variants and lentiCRISPRv2 (Addgene # #52961). $1x10^6$ cells were co-transfected with 1000 ng of an A3 variant (individual enzyme, not fused to nCas9) and 1000 ng of lentiCRISPRv2 (for Cas9 expression). 48 hours post-transfection, cells were harvested and lysed in 350 µl immunoprecipitation (IP) mild lysis buffer (50 mM Tris-HCl [pH 8], 150 mM NaCl, 0.8% NP-40, 10% glycerol, 1 mM phenylmethanesulfonyl fluoride solution [Sigma-Aldrich], and protease inhibitor cocktail set III [Calbiochem]). Samples were kept on ice for 20 minutes.

Lysates were subsequently cleared by centrifugation at maximum speed for 25 minutes at 4°C. 40 µl of supernatant were used as input samples, serving as a control. The supernatant was boiled at 95°C for 5 minutes with 15 µl loading buffer and stored at -20°C. The remaining lysate (IP sample) was incubated with 10 µl of anti-HA affinity matrix beads (Roche, Basel, Switzerland) at 4°C overnight with rotation. Prior to use, 10 µl of the beads had been resuspended in 100 µl gold buffer (50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 1% Triton X-100, 1 mM EDTA, 1 mM DTT, and protease inhibitor cocktail set III). The following day, samples were centrifuged at 3000 rpm at 4°C for 2 minutes. This was followed by three washing steps on ice using mild lysis buffer. After each wash the samples were centrifugated at 3000 rpm at 4°C for 2 minutes. Thereafter, the maximum amount of buffer was discarded without removing the beads. The remaining sample was then mixed with 10 µl of loading buffer and boiled for 5 minutes at 95°C.

IP and input samples were then simultaneously subjected to SDS-PAGE (10% gel) and immunoblotting, following the procedures as described above.

2.3. Statistical Analysis

All experiments were conducted independently in triplicate and the data are presented as the mean with standard deviation (SD) in all bar charts. Statistical significance was assessed using the unpaired Student’s t-test and one-way ANOVA with GraphPad Prism version 9.0.0 (GraphPad Software, San Diego, CA, USA). P-values < 0.05 were considered statistically significant.

2.4. Graphical illustrations

Graphical illustrations were created using BioRender's free basic version and Microsoft PowerPoint.

3. Results

3.1. A fluorescence-based real time reporter assay for BE activity

Are A3C variants efficient deaminases as part of BE constructs? In order to screen A3C-based editing complexes a real-time reporter system that generated green fluorescence upon cytidine base editing (Fig. 7) was utilized. The reporter system was based on three different expression plasmids that were co-transfected in 293T cells. The reporter plasmid carried a specific target site for the BE construct to bind and edit. The specific sgRNA ensured target recognition and binding of the base editor. Once the base editor successfully deaminated the target dC, a green fluorescent signal was emitted. Fluoresecent signals were detected either through fluorescence microscopy or through flowcytometry. In order to obtain a negative control, the base editors and the reporter plasmids were co-transfected with a NS sgRNA. While the NS sgRNA is still able to interact with the base editor, it does not match any sequence of the reporter plasmid. Therefore, it cannot guide the base editor to the editing target. The two established reporter plasmids, L202S and L138S carry, amongst others, genes for eGFP and mCherry). The eGFP sequence of both constructs contains a single T-to-C mutation at different positions that abolished eGFP fluorescence. The respective point-mutation represents a potential A3 deamination site. The L138S and L202S constructs introduced a leucine-to-serine mutation at positions 138 and 202, respectively. A mCherry gene, linked via a "self-cleaving" T2A peptide, served as a transfection or transduction control. Successfully transfected or transduced cells showed red fluorescence do to the mCherry expression. Green fluorescence was only displayed upon successful BE which restored the fluorescence-ablating mutation within the eGFP sequence. BE efficiency then was assessed by comparing double-positive (eGFP- and mCherry-positive) with single-positive (mCherry-only) cells (negative control) (Martin et al., 2019).

A panel of six A3-Cas9n-UGI constructs was tested for base editing efficiency. A3Ai, an established base editor, served as a positive control (Barka et al., 2022; Coelho et al., 2018; Gehrke et al., 2018; Wang et al., 2018). Five newly constructed base editors carried human A3C (huA3C) and four A3C variants. A3C S188I, A3C WE.RK, A3C CH2, and smmA3C-like were selected for this study due to increased catalytic activity (Adolph et

al., 2017; Jaguva Vasudevan et al., 2020; Wittkopp et al., 2016). The base editors were first subjected to episomal and then chromosomal BE activity screening experiments.

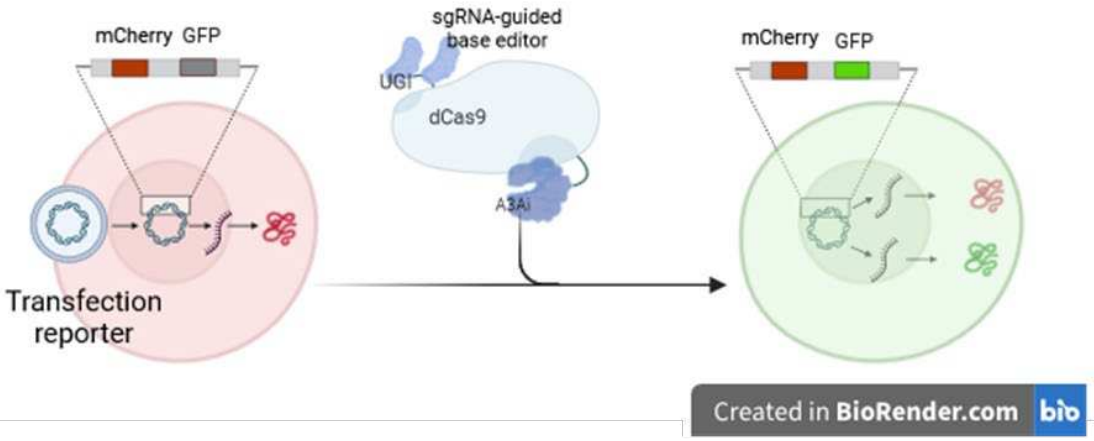

Figure 6. Schematic representation of the A3 base editor and eGFP reporter system. On the left side, only mCherry is expressed, eGFP expression is ablated due to a point mutation. The base editor, guided by a specific sgRNA, then induces deamination at the target site, restoring green fluorescence (right side).

3.2. BE activity screening of five A3C base editors

Transient transfection in 293T cells revealed that the L138S reporter yielded higher GFP restoration than the L202S reporter (Fig. 7a, c). As expected, A3Ai showed the highest editing efficiency. The results confirmed that huA3C does not display BE activity. Interestingly, however, the A3C CH2 editosome complex achieved comparable BE results to A3Ai for the L138S reporter (53.3% vs. 55.3% eGFP-positive cells, respectively). The other A3C base editor variants only yielded low editing frequencies. Base editor protein expression was confirmed via immunoblot analysis (Fig. 8a, c). To complement flowcytometry results, BE was also monitored through fluorescence microscopy. Representative fluorescence microscopy images of A3C CH2-mediated BE are given in Fig. 7b, d.

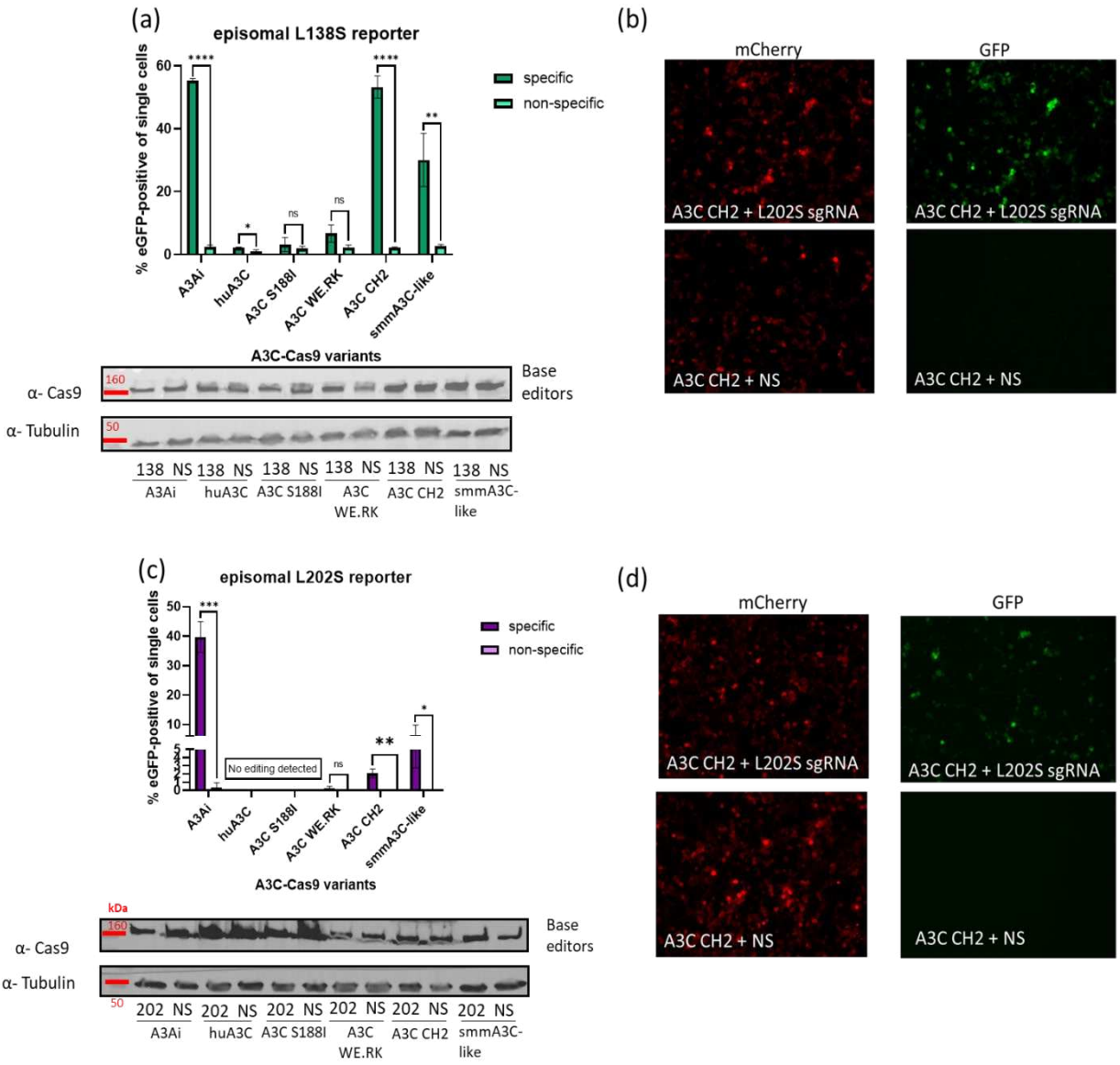

Figure 7. Base editing efficiency in episomal reporter assays (b) Flow cytometry analysis quantifying editing frequencies of A3Ai- and A3C variants-based editosomes using the L138S reporter in 293T cells (mean ± standard deviation). Protein expression of the base editors was confirmed via immunoblot analysis of cell lysates, with tubulin as a loading control ("α" represents anti). (c) Representative fluorescence microscopy images of 293T cells co-transfected with base editors, L138S reporter plasmids, and either a specific or non-specific (NS) sgRNA. (d-e) Base editing results for the L202S reporter, analyzed as in panels b and c.

3.3. Chromosomal BE experiments

The base editors were then tested for editing efficiency of chromosomal DNA. A pool of 293T cells constitutively expressing either the L202S or L138S reporter was generated through lentiviral vectors. Successful transduction led to the expression of mCherry and

positive cells were retrieved through fluorescent based cell sorting and cultured. These reporter cells were then transiently co-transfected with plasmids for base editors and appropriate sgRNA expression. Editing frequencies were lower than in the previous activity screening. However, A3C CH2 still exhibited comparable BE efficiency to A3Ai in the L138S reporter cell pool (28.2% vs. 33.1% eGFP-positive cells) (Fig. 8a). Consistently with episomal BE results, lower editing frequencies were observed in the L202S reporter cells (Fig. 8b). Immunoblot analysis confirmed BE expression (Fig. 8a, b).

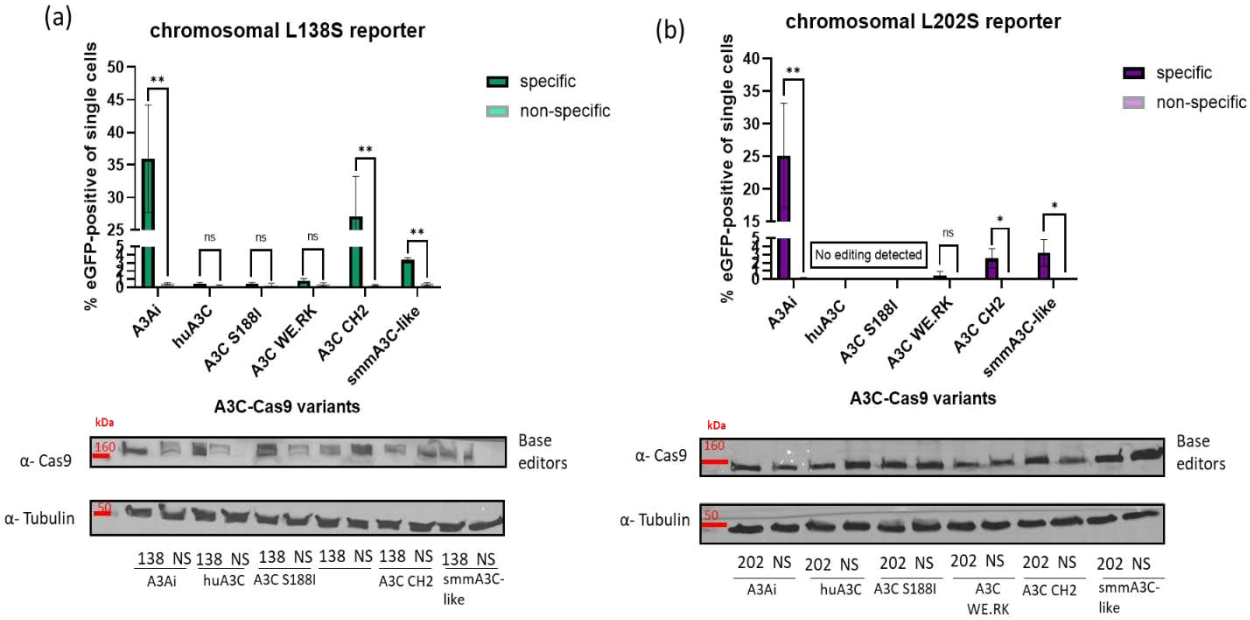

Figure 8. Base editing results for chromosomal reporters. (a-b) Flow cytometry analysis of base editing efficiencies in L138S- and L202S-eGFP reporter-transduced 293T cells (mean ± standard deviation). Cell pools were transiently transfected with the base editor panel and the corresponding sgRNA. Protein expression of the editing complexes was verified by immunoblot analysis, with tubulin serving as loading control. “NS” stands for non-specific sgRNA, and “α” represents anti**.**

3.4. All tested A3C based editing complexes interact similarly with Cas9

Based on the results presented above, the question surged why the A3C CH2 base editor yielded high editing frequencies, while the huA3C based editosome did not. First, to assess potential differences of subcellular base editor protein expression, confocal microscopy was conducted. The editing constructs all share the same backbone, which carries a nuclear localization sequence (NLS) downstream of the UGI. In contrast to

expectations, all tested base editors were almost exclusively expressed in the cytoplasm (Fig. 9). Interestingly, no notable differences in subcellular protein localization among active and inactive A3 editing constructs was observed.

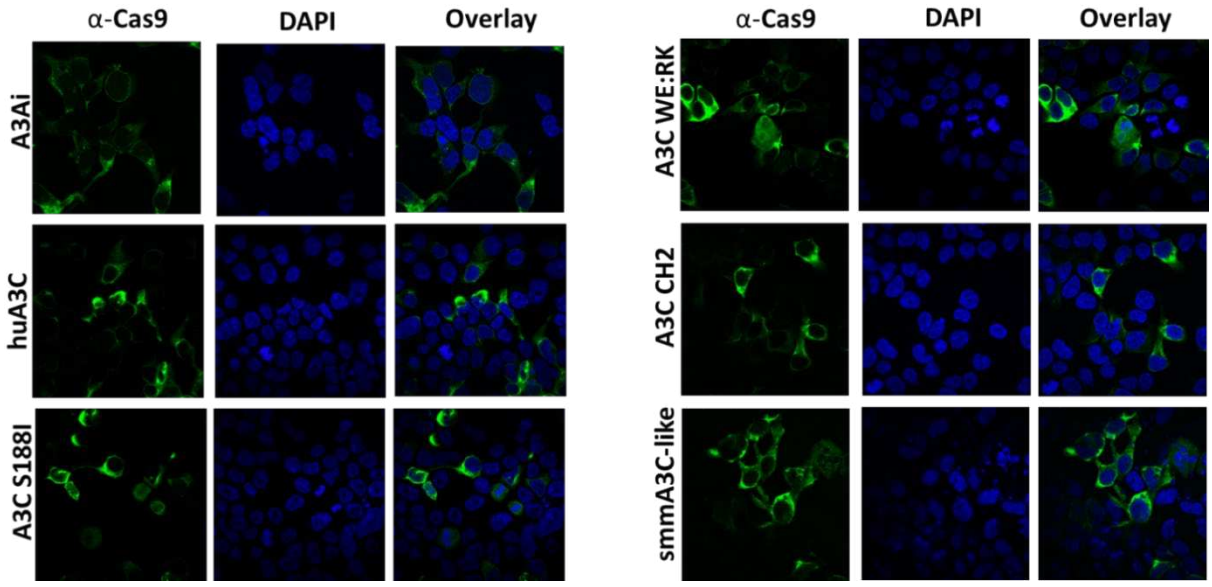

Figure 9. Localization of base editors within the cell. Confocal microscopy images of 293T cells transfected with base editors. A Cas9 antibody was used to locate protein expression. "α" represents anti.

To analyze A3C-nCas9 interactions, expression plasmids for A3A or A3C variants with Cas9 were co-transfected. The proteins were expressed separately and were not fused to each other. A3 proteins were expressed as HA-taged versions and detected by anti-HA staining. This ensured that observed interactions were independent of the BE activity itself. Immunoprecipitation (IP) assays using an anti-Cas9 antibody revealed that nCas9 exhibited comparable binding affinity across all tested A3 variants (Fig. 10a). This suggests that differences in nCas9 binding affinity do not attribute to the varying BE efficiencies among the tested base editors.

To assess native protein folding, I tested the sensitivity of A3C-based editing constructs to VIF-induced degradation. Lentiviral VIFs target certain A3 proteins for proteasomal degradation (Cadima-Couto et al., 2011; Kitamura et al., 2012). However, in the absence of a Vif recognition region or the masking of this region due to protein misfolding, the VIF cannot bind to the A3 protein. Therefore, in either of these cases the A3 protein will not be degraded and remain intact. To assess the base editors' sensitivity to VIF-induced degradation, BE complexes were co-transfected with the Simian Immunodeficiency Virus from the African green monkey (SIVagm) VIF expression plasmid. Immunoblot analysis

using an anti-Cas9 antibody showed that all tested A3C base editors were sensitive to SIVagm VIF. Importantly, both huA3C and A3C variants were degraded to same extent by VIF. As anticipated, A3A remained unaffected (Fig. 10b) (Aguiar et al., 2008).

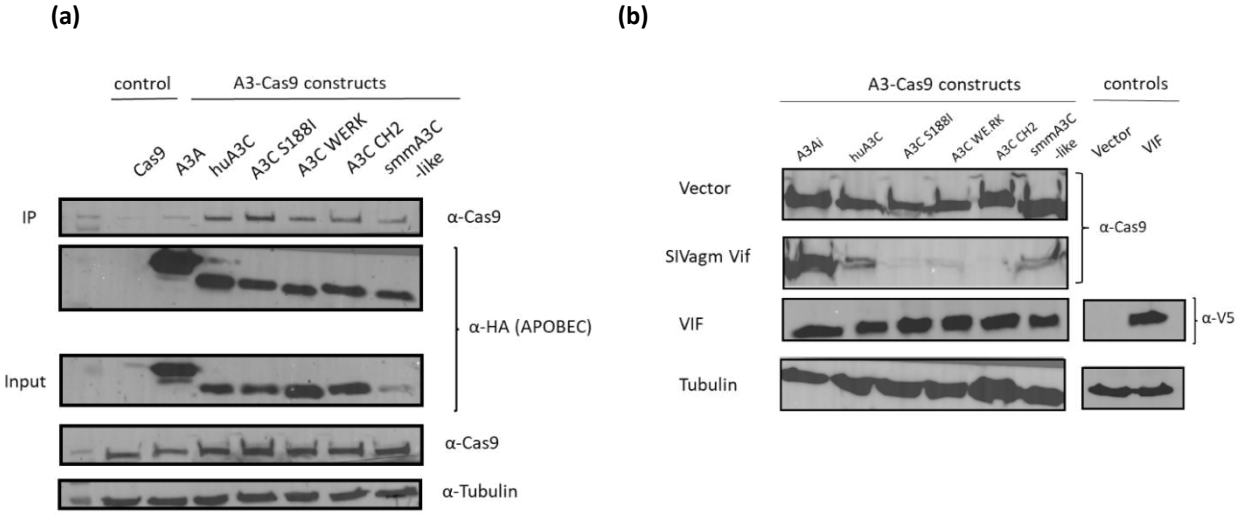

Figure 10. Interactivity of base editor proteins. (a) Interaction of A3 variants with Cas9. Immunoprecipitation of cell lysates from 293T cells co-transfected with A3 variants and nCas9 separately. Tubulin was used as a loading control, and "α" represents anti. (b) SIVagm VIF-mediated degradation of A3C-Cas9 base editors. 293T cells were co-transfected with the panel of base editor and SIVagm VIF. A3 protein expression was then determined through immunoblot analysis of cell lysates. Tubulin served as a loading control, with "α" indicating anti.

3.5. Loop 1 and Alpha-Helix 1 residues regulate A3C BE capacity

When aligning the sequences of huA3C and A3C CH2, 95% of amino acid residues were found to be identical. The first 36 amino acid residues of A3C CH2 are derived from the smmA3C-like protein, the remaining amino acids are derived from huA3C. Within the N-terminal region, huA3C and A3C CH2 merely differ in seven residues. These residues are part of the αHelix1 and loop 1 of the folded protein (Fig. 11a).

In A3 proteins, loop 1 is one of the least conserved regions and is known to influence substrate binding and enzymatic activity (Hou et al., 2021; Maiti et al., 2021; Ziegler et al., 2019). In addition, loop 1 was found to be crucial for A3C's antiviral activity against HIV-1 (Jaguva Vasudevan et al., 2020).

Suspecting that the differences in amino acid residues may account for A3C CH2 high BE activity in comparison to huA3C, new A3C variants with mutations to these two regions

were generated. Newly obtained base editors were then subjected to L138S reporter system to screen for BE activity.

The first approach was to create variants with a reciprocal exchange of αHelix1 between huA3C and A3C CH2. Flowcytometry results for the new variant huA3C.αHelix1$_{CH2}$ showed a loss-of function with a reduction of BE frequency from 53.3% to 0.4%. (Fig. 11b). Interestingly, the replacement of αHelix1 of huA3C with the αHelix1 of A3C CH2 did not lead to an increase of BE activity of huA3C.αHelix1$_{CH2.}$ Quite to the contrary, the already very low BE efficiency was completely diminished (Fig. 11b). Given these results, it was concluded that the αHelix1 of A3C CH2 cannot solely be responsible for the variant's high BE activity. To ensure that the loss of function was not due to a lack of protein expression, immunoblot analysis was performed on the newly obtained variants (Fig. 11c).

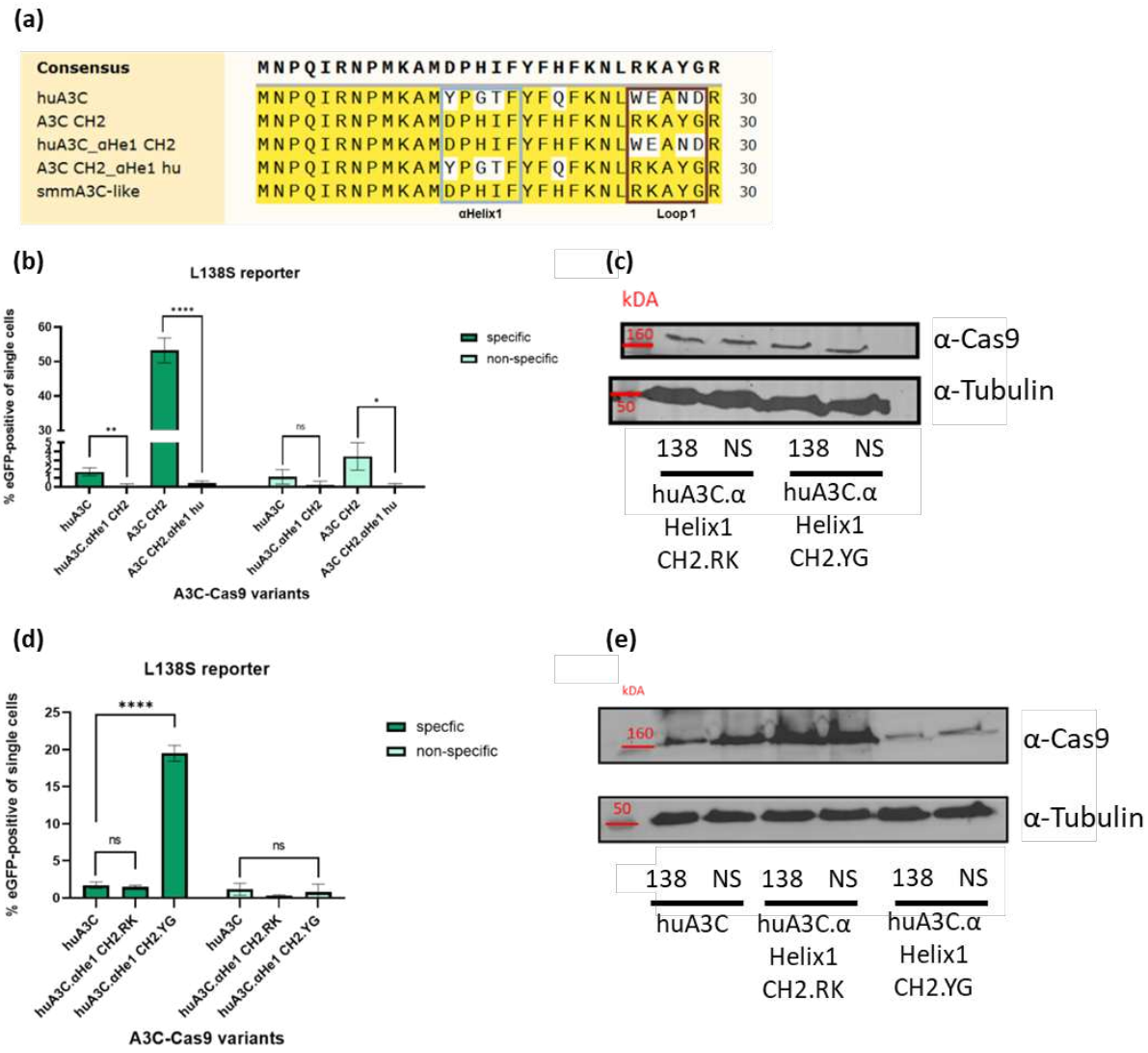

Figure 11. 3C base editor variants derived from human A3C and A3C CH2. (a) Amino acid alignment of the first 30 residues of the A3C variants, highlighting αHelix1 (blue outline) and Loop1 (red outline). (b) Flow cytometry analysis of newly developed base editors, comparing editing frequencies to prior episomal base editing experiments (mean ± standard deviation). 293T cells were co-transfected with base editors, L138S reporter plasmids, and either specific single guide RNA (sgRNA) or non-specific sgRNA (NS). Editing efficiencies of the original base editor and its derived variants were then compared. (c) Immunoblot analysis confirming protein expression, with tubulin serving as loading control. "α" represents anti. (d) Flow cytometry results for the additional base editor variants, performed as in panel. Editing efficiencies were compared to those of huA3C.αHelix1 CH2 (panel c) (mean ± standard deviation). (d) Immunoblot analysis confirming editosome protein expression, conducted as in panel c.

Therefore, the loop 1 in huA3C.αHelix1$_{CH2}$ became the region of interest for further variants. Loop 1 consists of five amino acid residues. A pairwise exchange of the first and last two residues was between huA3C and A3C CH2 led to two new variants. In variant huA3C.αHelix1$_{CH2}$RK, the first two amino acids of loop 1 ^{25}WE^{26} were mutated to ^{25}RK^{26}. To obtain the other new base editor, huA3C.αHelix1$_{CH2.}$YG, loop 1 residues ^{28}ND^{29} were mutated to ^{28}YG^{29}. The new base editors were then subjected to the L138S reporter assay. The BE activity was compared to previously obtained results for the huA3C editosome (Fig. 11d). The huA3C.αHelix1$_{CH2}$RK base editor did not show a statistically significant change of BE activity. However, the huA3C.αHelix1$_{CH2}$YG variant yielded 19.5% BE frequency, a highly significant increase when compared to huA3C (BE activity of 1.7%). This gain of function strongly indicated that the precise combination of αHelix1 of A3C CH2 and the Loop 1 residues ^{28}YG^{29} is highly important for A3C CH2's activity as base editor. Protein expression was ensured through immunoblot analysis (Fig. 9e). Remarkedly, the huA3C.αHelix1$_{CH2}$YG base editor showed lower protein expression compared to huA3C and huA3C.αHelix1$_{CH2}$RK. Given the variant's efficiency as base editor, it can be assumed that in spite of low expression, huA3C.αHelix1$_{CH2}$YG can unfold its enzymatic activity.

4. Discussion

Compared to other human A3 enzymes such as A3A or A3B, only a few studies have explored the use of A3C as part of a CBE complex. The limited available data strongly suggests that A3C does not exhibit BE efficiency (Jin et al., 2020; Liu et al., 2019; Martin et al., 2019). The results of this study confirm that huA3C is not an efficient base editor. However, multiple variants of huA3C were introduced as part of the BE complex during the project. The selection of these variants was based on previous studies showing increased catalytic and antiviral activity of the A3C variants. Wittkopp et al. identified a single nucleotide polymorphism in A3C that increased its ability to restrict lentiviruses, suggesting that A3C S188I may have enhanced catalytic activity. A study conducted and published by our research group demonstrated that Loop 1 of A3C plays a crucial role in regulating its antiviral activity. A3C WE.RK is a variant with mutations to the Loop 1 region. smmA3C-like is a variant derived from sooty mangabey and has enhanced enzymatic and antiviral properties. Vasudevan et al. also created chimera from huA3C and smmA3C-like out of which especially A3C CH2 displayed high enzymatic activity (Jaguva Vasudevan et al., 2020; Wittkopp et al., 2016).

Given these findings, we included these specific A3C variants for our BE complexes. We hypothesized that these characteristics would translate to improved BE efficiency. An editing construct carrying huA3C was also created to both test for BE utility and allow for a comparison with A3C variants. The established A3A base editor served as positive control for our experiments (St Martin et al., 2018).

Using the mCherry-eGFP real-time reporter assay established in this work, the base editor A3C CH2 showed BE activity comparable to the A3A base editor. In contrast, A3C variants A3C S188I and A3C WE.RK despite of increased enzymatic activity, proved to be inefficient base editors. The variant smmA3C-like yielded higher editing frequencies than the inefficient variants, but was not up to par with the A3C CH2 base editor. Furthermore, we could confirm that huA3C does not show catalytic activity in context of BE.

These results were critically analyzed, and various theories surged to explain the high BE activity of A3C CH2 in particular. Initially, no apparent differences between active and inactive base editors could be identified. Confocal microscopy showed that both active

and inactive A3C-based base editors are expressed primarily in the cytoplasm. Suggesting that the subcellular localization of the editors is not a distinguishing factor between active and inactive versions. This ruled out the hypothesis that compartmentalization could affect BE efficiency in this study. Immunoprecipitation assays revealed that Cas9 interacts equally with A3A, huA3C, and A3C variants, suggesting no differences in protein interaction or binding affinity. This implies that the catalytic machinery involving Cas9 and the A3C base editors is functioning in a similar manner across the examined variants. Therefore, variations in BE efficiency are unlikely due to differences in protein interaction between nCas9 and the editing construct. To examine sensitivity to degradation, BE complexes were co-transfected with SIVagm VIF. The sensitivity to degradation can provide insights into potential misfolding or structural instability of the BE complexes. This could in turn affect the BE efficiency. The A3A base editor expression remained stable as A3A is not sensitive to SIVagm VIF-induced degradation. In contrast, all A3C-BE variants were degraded to the same extent (Cadima-Couto et al., 2011; Kitamura et al., 2012). While these results do not rule out protein misfolding completely, it can be assumed that variations in BE efficiency are more likely to stem from other molecular mechanisms.

R-loops play a crucial role in BE by influencing target accessibility and editing efficiency. R-loops are three-stranded nucleic acid structures consisting of an RNA-DNA hybrid and a displaced single-stranded DNA. R-loop formation is essential for exposing the target DNA strand for the deaminase to modify the targeted base. In the attempt to explain BE efficiency differences between A3C variants, the editing constructs were subjected to an adapted R-loop assay (Fig. 12). This experiment was conducted by Mirriam Nzivo, a current PhD student at the lab. Results showed that both A3A- and A3C CH2-based editing complexes are able to form R-loops. The huA3C base editor did not show R-loop formation suggesting that R-loop formation is a key determinant of BE efficiency in this study (Jiang & Doudna, 2017). The inability of huA3C-based editosome to form R-loops, compared to A3C CH2, may be due to differences in their structural or functional properties that affect interaction with nucleic acids. A3C CH2 may possess structural features that allow it to stabilize ssDNA regions required for R-loop formation, such as additional loops or helices that promote DNA unwinding. In contrast, the huA3C base

editor might lack these features or have configurations that prevent unwinding and R-loop formation. Another factor attributing to this significant difference could be a higher affinity of the A3C CH2 editosome for ssDNA. The huA3C base editor, on the other hand, may not interact as strongly with ssDNA making R-loop formation inefficient. The presence of cofactors like DNA-binding proteins might also influence R-loop formation, with A3C CH2 possibly being more efficient at recruiting these factors (Hegazy et al., 2020; Mazina et al., 2020; Petermann et al., 2022). Thus, numerous factors may account for huA3C and A3C CH2 base editors' distinct abilities to form R-loops. Further research focused exploring these theories can contribute to a better understanding not only of the A3C variants, but more importantly of CBE mechanisms.

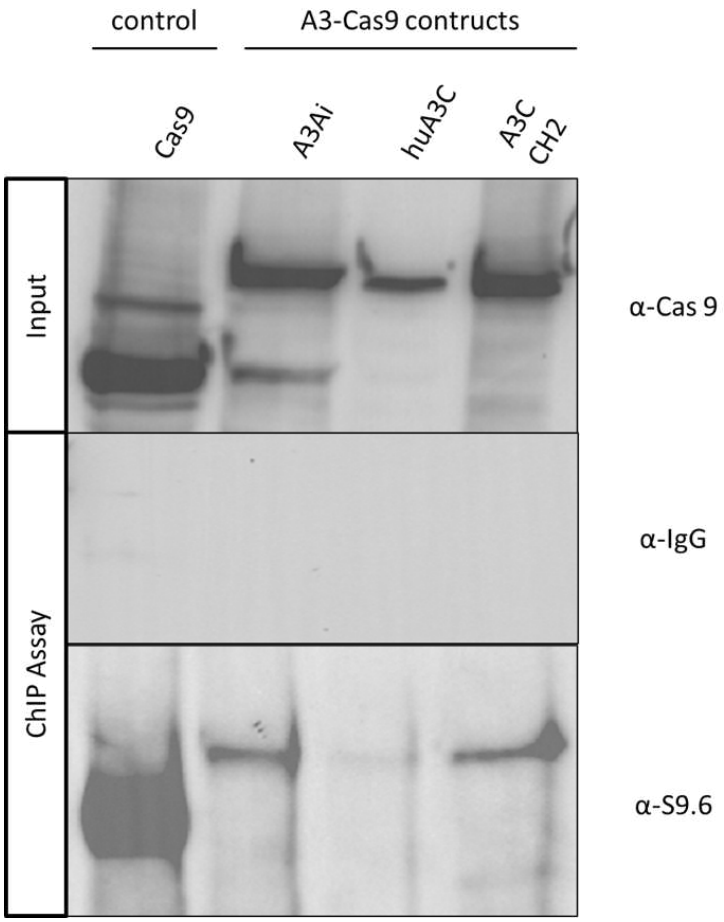

Figure 12. A3 base editor-dependent R-loop formation. Modified ChIP assay performed on 293T cells transfected with A3Ai, huA3C, A3C CH2 editing complexes, or Cas9 individually. Immunoblot analysis of eluted proteins was conducted, with α-Cas9 used to detect protein expression and α-IgG serving as a negative control. "α" stands for anti.

Another approach to further characterize the A3C variants in BE encompassed the creating of further A3C variants to identify the key amino acid residues contributing to

A3C CH2's high BE efficiently. The newly generated A3C base editor variants carry mutations in the Loop1 and αHelix1 regions. Within the A3 protein family, these two regions are considered critical for substrate binding and enzymatic activity (Harjes et al., 2009; Jaguva Vasudevan et al., 2020; Kim et al., 2023; Maiti et al., 2021; Maiti et al., 2020; Ziegler et al., 2019). These new base editors—huA3C.αHe1 CH2, A3C CH2.αHe1 hu, huA3C.αHe1 CH2.RK, and huA3C.αHe1 CH2.YG—were subjected to the same real-time reporter assay as the earlier A3C variants. Interestingly, only huA3C.αHe1 CH2.YG, derived from huA3C with A3C CH2 αHelix1 and Loop1 amino acid substitutions ^{28}YG^{29}, showed a functional enhancement as a base editor. Therefore, we hypothesize that the combination of these two modified residues significantly contributes to the high efficiency of the A3C CH2 base editor. This exact sequence may be essential for stabilizing ssDNA targets (Lin et al., 2021; Wang et al., 2017).Characterizing the key amino acid residues may provide a foundation for optimizing other base editors by targeting key regions that enhance their editing capabilities while minimizing off-target effects. The development of new base editors utilizing different deaminases is fundamental for the continuous improvement of BE (Huang et al., 2021).

This study focused on A3C variants as base editors for ssDNA. An open question is whether the A3C CH2 base editor can also edit RNA. Previous studies have shown that A3A and A3G have the ability to bind to RNA and deaminate it. A3A binds RNA with an affinity comparable to that of ssDNA and has a broader range of RNA target sites than A3G (Kim et al., 2023; Sharma et al., 2016; Sharma et al., 2015; Solomon et al., 2019). Potential effects of A3C CH2 on RNA were not within this study, as it would have exceeded its research scope. Nonetheless, this provides a starting point for further research. A systematic investigation into its RNA-editing capacity would require targeted biochemical and molecular analyses to assess RNA binding and deamination activity. Investigating the RNA-editing potential of A3C CH2 could provide valuable insights into the structural and functional factors that influence A3-inducted RNA deamination, thus contributing to the development of new RNA-editing technologies.

Another open question was why active variants exhibited BE efficiency primarily in the L138S reporter system. Editing frequencies were significantly lower when tested with the L202S reporter plasmid. A3 enzymes typically show a strong preference for

deaminating TC dinucleotides. The deamination site in both plasmids had the TC sequence. However, the mere presence of a TC motif does not guarantee ssDNA deamination (Shi et al., 2017). Brown identified A3C's preferred sequence context, extending 3 nucleotides upstream and downstream of the deamination site. Based on this motif, the L202S reporter plasmid would be expected to yield higher A3C-dependent BE activity. This plasmid features A3C's preferred target sequence and lacks disfavored multiple G and C nucleotides at positions +1, +2, and +3. Despite this theoretical advantage, our A3C base editors exhibited greater editing activity with the L138S reporter plasmid. This discrepancy highlights the complexity of A3C-mediated deamination, which is influenced not only by sequence motifs but also by factors such as ssDNA structural accessibility, intrinsic enzyme properties, and the broader genomic context (Brown, 2024). Additionally, A3A compared to A3C has a broader target sequence specificity in ssDNA due to more flexible substrate binding and less restrictive sequence preferences (Chan et al., 2015; Hou et al., 2021; McDaniel et al., 2020). This may explain why the A3A base editor showed activity in both reporter systems.

Compared to conventional CRISPR/Cas9 systems, base editors are less prone to undesirable mutations, such as insertions and deletions. Nevertheless, challenges remain in base editing despite ongoing efforts to improve the technology. The detailed characterization of base editors, as conducted in this study, can serve as a foundation for developing new base editors with higher efficiency and reduced off-target effects (Naeem et al., 2020; Rees et al., 2017; Slesarenko et al., 2022).

This study demonstrates the potential of A3C variants as base editors, with A3C CH2 showing BE efficiency comparable with the established A3A base editor. Results suggest that structural elements of A3C CH2, such as its ability to form R-loops, may contribute to its enhanced editing efficiency. In contrast, other A3C variants, in spite of high enzymatic activity, did not exhibit similar capacity in the context of BE (Jaguva Vasudevan et al., 2020). This emphasizes the complexity and highly variable factors to be considered for the development of new base editors.

The results of this study underline the importance of identifying molecular factors that influence BE efficiency, particularly those related to R-loop formation and DNA targeting. This may contribute to address issues such as off-target DNA and RNA editing (Park &

Beal, 2019). Additionally, exploring RNA-editing capabilities of various base editors could expand the potential application in RNA-based gene therapy. Further optimizing BE technologies and understanding the key determinants of their activity is essential for the continuous improvement of this powerful editing tool (Budzko et al., 2023; Thuronyi et al., 2019). BE represents a safer alternative to CRISPR/Cas9 gene therapy for addressing genetic disorders caused by point mutations (Newby & Liu, 2021; Newby et al., 2021).

5. References

Adolph, M. B., Ara, A., Feng, Y., Wittkopp, C. J., Emerman, M., Fraser, J. S., & Chelico, L. (2017). Cytidine deaminase efficiency of the lentiviral viral restriction factor APOBEC3C correlates with dimerization. *Nucleic Acids Res*, *45*(6), 3378-3394. https://doi.org/10.1093/nar/gkx066

Aguiar, R. S., Lovsin, N., Tanuri, A., & Peterlin, B. M. (2008). Vpr.A3A chimera inhibits HIV replication. *J Biol Chem*, *283*(5), 2518-2525. https://doi.org/10.1074/jbc.M706436200

Attia, E. F., Maleche-Obimbo, E., Ellington, L. E., & North, C. M. (2025). Pulmonary Immunocompromise in Human Immunodeficiency Virus Disease. *Clin Chest Med*, *46*(1), 185-201. https://doi.org/10.1016/j.ccm.2024.10.014

Barka, A., Berríos, K. N., Bailer, P., Schutsky, E. K., Wang, T., & Kohli, R. M. (2022). The Base-Editing Enzyme APOBEC3A Catalyzes Cytosine Deamination in RNA with Low Proficiency and High Selectivity. *ACS Chem Biol*, *17*(3), 629-636. https://doi.org/10.1021/acschembio.1c00919

Biolatti, M., Gugliesi, F., Dell'Oste, V., & Landolfo, S. (2018). Modulation of the innate immune response by human cytomegalovirus. *Infect Genet Evol*, *64*, 105-114. https://doi.org/10.1016/j.meegid.2018.06.025

Bishop, K. N., Holmes, R. K., Sheehy, A. M., Davidson, N. O., Cho, S. J., & Malim, M. H. (2004). Cytidine deamination of retroviral DNA by diverse APOBEC proteins. *Curr Biol*, *14*(15), 1392-1396. https://doi.org/10.1016/j.cub.2004.06.057

Brown, G. W. (2024). The cytidine deaminase APOBEC3C has unique sequence and genome feature preferences. *Genetics*, *227*(4). https://doi.org/10.1093/genetics/iyae092

Budzko, L., Hoffa-Sobiech, K., Jackowiak, P., & Figlerowicz, M. (2023). Engineered deaminases as a key component of DNA and RNA editing tools. *Mol Ther Nucleic Acids*, *34*, 102062. https://doi.org/10.1016/j.omtn.2023.102062

Cadima-Couto, I., Saraiva, N., Santos, A. C., & Goncalves, J. (2011). HIV-1 Vif interaction with APOBEC3 deaminases and its characterization by a new sensitive assay. *J Neuroimmune Pharmacol*, *6*(2), 296-307. https://doi.org/10.1007/s11481-011-9258-7

Chan, K., Roberts, S. A., Klimczak, L. J., Sterling, J. F., Saini, N., Malc, E. P., Kim, J., Kwiatkowski, D. J., Fargo, D. C., Mieczkowski, P. A., Getz, G., & Gordenin, D. A. (2015). An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers. *Nat Genet*, *47*(9), 1067-1072. https://doi.org/10.1038/ng.3378

Chen, X., Du, J., Yun, S., Xue, C., Yao, Y., & Rao, S. (2024). Recent advances in CRISPR-Cas9-based genome insertion technologies. *Mol Ther Nucleic Acids*, *35*(1), 102138. https://doi.org/10.1016/j.omtn.2024.102138

Coelho, M. A., Li, S., Pane, L. S., Firth, M., Ciotta, G., Wrigley, J. D., Cuomo, M. E., Maresca, M., & Taylor, B. J. M. (2018). BE-FLARE: a fluorescent reporter of base editing activity reveals editing characteristics of APOBEC3A and APOBEC3B. *BMC Biol*, *16*(1), 150. https://doi.org/10.1186/s12915-018-0617-1

Cowan, F. M., Shahmanesh, M., Revill, P. A., Busza, J., Sibanda, E. L., Chabata, S. T., Chimbindi, N., Choola, T., Mugurungi, O., Hargreaves, J. R., & Phillips, A. N.

(2025). Preventing HIV in women in Africa. *Nat Med*. https://doi.org/10.1038/s41591-025-03535-8

Dennis, M., Hurley, A., Bray, N., Cordero, C., Ilagan, J., Mertz, T. M., & Roberts, S. A. (2024). Her2 amplification, Rel-A, and Bach1 can influence APOBEC3A expression in breast cancer cells. *PLoS Genet*, *20*(5), e1011293. https://doi.org/10.1371/journal.pgen.1011293

Ford, N., Kassanjee, R., Stelzle, D., Jarvis, J. N., Sued, O., Perrin, G., Doherty, M., & Rangaraj, A. (2025). Global prevalence of advanced HIV disease in healthcare settings: a rapid review. *J Int AIDS Soc*, *28*(2), e26415. https://doi.org/10.1002/jia2.26415

Gehrke, J. M., Cervantes, O., Clement, M. K., Wu, Y., Zeng, J., Bauer, D. E., Pinello, L., & Joung, J. K. (2018). An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. *Nat Biotechnol*, *36*(10), 977-982. https://doi.org/10.1038/nbt.4199

Ghosn, J., Taiwo, B., Seedat, S., Autran, B., & Katlama, C. (2018). HIV. *Lancet*, *392*(10148), 685-697. https://doi.org/10.1016/s0140-6736(18)31311-4

Goila-Gaur, R., & Strebel, K. (2008). HIV-1 Vif, APOBEC, and intrinsic immunity. *Retrovirology*, *5*, 51. https://doi.org/10.1186/1742-4690-5-51

Gurrola, T. E., Effah, S. N., Sariyer, I. K., Dampier, W., Nonnemacher, M. R., & Wigdahl, B. (2024). Delivering CRISPR to the HIV-1 reservoirs. *Front Microbiol*, *15*, 1393974. https://doi.org/10.3389/fmicb.2024.1393974

Harjes, E., Gross, P. J., Chen, K. M., Lu, Y., Shindo, K., Nowarski, R., Gross, J. D., Kotler, M., Harris, R. S., & Matsuo, H. (2009). An extended structure of the APOBEC3G catalytic domain suggests a unique holoenzyme model. *J Mol Biol*, *389*(5), 819-832. https://doi.org/10.1016/j.jmb.2009.04.031

Harris, R. S., Bishop, K. N., Sheehy, A. M., Craig, H. M., Petersen-Mahrt, S. K., Watt, I. N., Neuberger, M. S., & Malim, M. H. (2003). DNA deamination mediates innate immunity to retroviral infection. *Cell*, *113*(6), 803-809. https://doi.org/10.1016/s0092-8674(03)00423-9

Hegazy, Y. A., Fernando, C. M., & Tran, E. J. (2020). The balancing act of R-loop biology: The good, the bad, and the ugly. *J Biol Chem*, *295*(4), 905-913. https://doi.org/10.1074/jbc.REV119.011353

Hou, S., Lee, J. M., Myint, W., Matsuo, H., Kurt Yilmaz, N., & Schiffer, C. A. (2021). Structural basis of substrate specificity in human cytidine deaminase family APOBEC3s. *J Biol Chem*, *297*(2), 100909. https://doi.org/10.1016/j.jbc.2021.100909

Hu, Z., Wang, Y., Liu, Q., Qiu, Y., Zhong, Z., Li, K., Li, W., Deng, Z., & Sun, Y. (2021). Improving the Precision of Base Editing by Bubble Hairpin Single Guide RNA. *mBio*, *12*(2). https://doi.org/10.1128/mBio.00342-21

Huang, T. P., Newby, G. A., & Liu, D. R. (2021). Precision genome editing using cytosine and adenine base editors in mammalian cells. *Nat Protoc*, *16*(2), 1089-1128. https://doi.org/10.1038/s41596-020-00450-9

Jaguva Vasudevan, A. A., Balakrishnan, K., Gertzen, C. G. W., Borveto, F., Zhang, Z., Sangwiman, A., Held, U., Kustermann, C., Banerjee, S., Schumann, G. G., Haussinger, D., Bravo, I. G., Gohlke, H., & Munk, C. (2020). Loop 1 of APOBEC3C Regulates its Antiviral Activity against HIV-1. *J Mol Biol*, *432*(23), 6200-6227. https://doi.org/10.1016/j.jmb.2020.10.014

Jeong, Y. K., Lee, S., Hwang, G. H., Hong, S. A., Park, S. E., Kim, J. S., Woo, J. S., & Bae, S. (2021). Adenine base editor engineering reduces editing of bystander cytosines. *Nat Biotechnol*, *39*(11), 1426-1433. https://doi.org/10.1038/s41587-021-00943-2

Jiang, F., & Doudna, J. A. (2017). CRISPR-Cas9 Structures and Mechanisms. *Annu Rev Biophys*, *46*, 505-529. https://doi.org/10.1146/annurev-biophys-062215-010822

Jin, S., Fei, H., Zhu, Z., Luo, Y., Liu, J., Gao, S., Zhang, F., Chen, Y. H., Wang, Y., & Gao, C. (2020). Rationally Designed APOBEC3B Cytosine Base Editors with Improved Specificity. *Mol Cell*, *79*(5), 728-740.e726. https://doi.org/10.1016/j.molcel.2020.07.005

Karn, J., & Stoltzfus, C. M. (2012). Transcriptional and posttranscriptional regulation of HIV-1 gene expression. *Cold Spring Harb Perspect Med*, *2*(2), a006916. https://doi.org/10.1101/cshperspect.a006916

Kim, K., Shi, A. B., Kelley, K., & Chen, X. S. (2023). Unraveling the Enzyme-Substrate Properties for APOBEC3A-Mediated RNA Editing. *J Mol Biol*, *435*(17), 168198. https://doi.org/10.1016/j.jmb.2023.168198

Kitamura, S., Ode, H., Nakashima, M., Imahashi, M., Naganawa, Y., Kurosawa, T., Yokomaku, Y., Yamane, T., Watanabe, N., Suzuki, A., Sugiura, W., & Iwatani, Y. (2012). The APOBEC3C crystal structure and the interface for HIV-1 Vif binding. *Nat Struct Mol Biol*, *19*(10), 1005-1010. https://doi.org/10.1038/nsmb.2378

Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., & Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, *533*(7603), 420-424. https://doi.org/10.1038/nature17946

Komor, A. C., Zhao, K. T., Packer, M. S., Gaudelli, N. M., Waterbury, A. L., Koblan, L. W., Kim, Y. B., Badran, A. H., & Liu, D. R. (2017). Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. *Sci Adv*, *3*(8), eaao4774. https://doi.org/10.1126/sciadv.aao4774

Kulpa, D. A., Paiardini, M., & Silvestri, G. (2025). Immune-mediated strategies to solving the HIV reservoir problem. *Nat Rev Immunol*. https://doi.org/10.1038/s41577-025-01136-7

Kwong, P. D., Wyatt, R., Robinson, J., Sweet, R. W., Sodroski, J., & Hendrickson, W. A. (1998). Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature*, *393*(6686), 648-659. https://doi.org/10.1038/31405

Langlois, M. A., Beale, R. C., Conticello, S. G., & Neuberger, M. S. (2005). Mutational comparison of the single-domained APOBEC3C and double-domained APOBEC3F/G anti-retroviral cytidine deaminases provides insight into their DNA target site specificities. *Nucleic Acids Res*, *33*(6), 1913-1923. https://doi.org/10.1093/nar/gki343

Lesbats, P., Engelman, A. N., & Cherepanov, P. (2016). Retroviral DNA Integration. *Chem Rev*, *116*(20), 12730-12757. https://doi.org/10.1021/acs.chemrev.6b00125

Lin, M., Malik, F. K., & Guo, J. T. (2021). A comparative study of protein-ssDNA interactions. *NAR Genom Bioinform*, *3*(1), lqab006. https://doi.org/10.1093/nargab/lqab006

Liu, H., Zhu, Y., Li, M., & Gu, Z. (2023). Precise genome editing with base editors. *Med Rev (2021)*, *3*(1), 75-84. https://doi.org/10.1515/mr-2022-0044

Liu, Z., Chen, S., Shan, H., Jia, Y., Chen, M., Song, Y., Lai, L., & Li, Z. (2020). Precise base editing with CC context-specificity using engineered human APOBEC3G-nCas9 fusions. *BMC Biol*, *18*(1), 111. https://doi.org/10.1186/s12915-020-00849-6

Liu, Z., Chen, S., Shan, H., Zhang, Q., Chen, M., Lai, L., & Li, Z. (2019). Efficient and precise base editing in rabbits using human APOBEC3A-nCas9 fusions. *Cell Discov*, *5*, 31. https://doi.org/10.1038/s41421-019-0099-5

Maiti, A., Hou, S., Schiffer, C. A., & Matsuo, H. (2021). Interactions of APOBEC3s with DNA and RNA. *Curr Opin Struct Biol*, *67*, 195-204. https://doi.org/10.1016/j.sbi.2020.12.004

Maiti, A., Myint, W., Delviks-Frankenberry, K. A., Hou, S., Kanai, T., Balachandran, V., Sierra Rodriguez, C., Tripathi, R., Kurt Yilmaz, N., Pathak, V. K., Schiffer, C. A., & Matsuo, H. (2020). Crystal Structure of a Soluble APOBEC3G Variant Suggests ssDNA to Bind in a Channel that Extends between the Two Domains. *J Mol Biol*, *432*(23), 6042-6060. https://doi.org/10.1016/j.jmb.2020.10.020

Martin, A. S., Salamango, D. J., Serebrenik, A. A., Shaban, N. M., Brown, W. L., & Harris, R. S. (2019). A panel of eGFP reporters for single base editing by APOBEC-Cas9 editosome complexes. *Sci Rep*, *9*(1), 497. https://doi.org/10.1038/s41598-018-36739-9

Mazina, O. M., Somarowthu, S., Kadyrova, L. Y., Baranovskiy, A. G., Tahirov, T. H., Kadyrov, F. A., & Mazin, A. V. (2020). Replication protein A binds RNA and promotes R-loop formation. *J Biol Chem*, *295*(41), 14203-14213. https://doi.org/10.1074/jbc.RA120.013812

McDaniel, Y. Z., Wang, D., Love, R. P., Adolph, M. B., Mohammadzadeh, N., Chelico, L., & Mansky, L. M. (2020). Deamination hotspots among APOBEC3 family members are defined by both target site sequence context and ssDNA secondary structure. *Nucleic Acids Res*, *48*(3), 1353-1371. https://doi.org/10.1093/nar/gkz1164

Mishra, R., Joshi, R. K., & Zhao, K. (2020). Base editing in crops: current advances, limitations and future implications. *Plant Biotechnol J*, *18*(1), 20-31. https://doi.org/10.1111/pbi.13225

Modenini, G., Abondio, P., & Boattini, A. (2022). The coevolution between APOBEC3 and retrotransposons in primates. *Mob DNA*, *13*(1), 27. https://doi.org/10.1186/s13100-022-00283-1

Naeem, M., Majeed, S., Hoque, M. Z., & Ahmad, I. (2020). Latest Developed Strategies to Minimize the Off-Target Effects in CRISPR-Cas-Mediated Genome Editing. *Cells*, *9*(7). https://doi.org/10.3390/cells9071608

Newby, G. A., & Liu, D. R. (2021). In vivo somatic cell base editing and prime editing. *Mol Ther*, *29*(11), 3107-3124. https://doi.org/10.1016/j.ymthe.2021.09.002

Newby, G. A., Yen, J. S., Woodard, K. J., Mayuranathan, T., Lazzarotto, C. R., Li, Y., Sheppard-Tillman, H., Porter, S. N., Yao, Y., Mayberry, K., Everette, K. A., Jang, Y., Podracky, C. J., Thaman, E., Lechauve, C., Sharma, A., Henderson, J. M., Richter, M. F., Zhao, K. T., . . . Liu, D. R. (2021). Base editing of haematopoietic stem cells rescues sickle cell disease in mice. *Nature*, *595*(7866), 295-302. https://doi.org/10.1038/s41586-021-03609-w

Okamoto, E. E., Anam, F. R., Batiste, S., Dukashe, M., Castellanos, E., Poonkasetwattana, M., & Richman, B. (2024). Ending AIDS as a public health threat: the imperative for clear messaging on U=U, viral suppression, and zero risk. *Lancet HIV*, *11*(11), e783-e790. https://doi.org/10.1016/s2352-3018(24)00241-8

Park, S., & Beal, P. A. (2019). Off-Target Editing by CRISPR-Guided DNA Base Editors. *Biochemistry*, *58*(36), 3727-3734. https://doi.org/10.1021/acs.biochem.9b00573

Petermann, E., Lan, L., & Zou, L. (2022). Sources, resolution and physiological relevance of R-loops and RNA-DNA hybrids. *Nat Rev Mol Cell Biol*, *23*(8), 521-540. https://doi.org/10.1038/s41580-022-00474-x

Poliektov, N. E., & Badell, M. L. (2023). Antiretroviral Options and Treatment Decisions During Pregnancy. *Paediatr Drugs*, *25*(3), 267-282. https://doi.org/10.1007/s40272-023-00559-w

Porto, E. M., Komor, A. C., Slaymaker, I. M., & Yeo, G. W. (2020). Base editing: advances and therapeutic opportunities. *Nat Rev Drug Discov*, *19*(12), 839-859. https://doi.org/10.1038/s41573-020-0084-6

Rashid, F., Zaongo, S. D., Iqbal, H., Harypursat, V., Song, F., & Chen, Y. (2024). Interactions between HIV proteins and host restriction factors: implications for potential therapeutic intervention in HIV infection. *Front Immunol*, *15*, 1390650. https://doi.org/10.3389/fimmu.2024.1390650

Rees, H. A., Komor, A. C., Yeh, W. H., Caetano-Lopes, J., Warman, M., Edge, A. S. B., & Liu, D. R. (2017). Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. *Nat Commun*, *8*, 15790. https://doi.org/10.1038/ncomms15790

Rees, H. A., & Liu, D. R. (2018). Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat Rev Genet*, *19*(12), 770-788. https://doi.org/10.1038/s41576-018-0059-1

Ren, C. Y., Liu, Y. S., He, Y. S., Zhang, L. P., Rao, J. H., Rao, Y., & Chen, J. H. (2024). Engineered CBEs based on Macaca fascicularis A3A with improved properties for precise genome editing. *Cell Rep*, *43*(3), 113878. https://doi.org/10.1016/j.celrep.2024.113878

Salter, J. D., Bennett, R. P., & Smith, H. C. (2016). The APOBEC Protein Family: United by Structure, Divergent in Function. *Trends Biochem Sci*, *41*(7), 578-594. https://doi.org/10.1016/j.tibs.2016.05.001

Schemelev, A. N., Davydenko, V. S., Ostankova, Y. V., Reingardt, D. E., Serikova, E. N., Zueva, E. B., & Totolian, A. A. (2024). Involvement of Human Cellular Proteins and Structures in Realization of the HIV Life Cycle: A Comprehensive Review, 2024. *Viruses*, *16*(11). https://doi.org/10.3390/v16111682

Sharma, S., Patnaik, S. K., Taggart, R. T., & Baysal, B. E. (2016). The double-domain cytidine deaminase APOBEC3G is a cellular site-specific RNA editing enzyme. *Sci Rep*, *6*, 39100. https://doi.org/10.1038/srep39100

Sharma, S., Patnaik, S. K., Taggart, R. T., Kannisto, E. D., Enriquez, S. M., Gollnick, P., & Baysal, B. E. (2015). APOBEC3A cytidine deaminase induces RNA editing in monocytes and macrophages. *Nat Commun*, *6*, 6881. https://doi.org/10.1038/ncomms7881

Sharp, P. M., & Hahn, B. H. (2011). Origins of HIV and the AIDS pandemic. *Cold Spring Harb Perspect Med*, *1*(1), a006841. https://doi.org/10.1101/cshperspect.a006841

Shi, K., Carpenter, M. A., Banerjee, S., Shaban, N. M., Kurahashi, K., Salamango, D. J., McCann, J. L., Starrett, G. J., Duffy, J. V., Demir, O., Amaro, R. E., Harki, D. A., Harris, R. S., & Aihara, H. (2017). Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B. *Nat Struct Mol Biol*, *24*(2), 131-139. https://doi.org/10.1038/nsmb.3344

Slesarenko, Y. S., Lavrov, A. V., & Smirnikhina, S. A. (2022). Off-target effects of base editors: what we know and how we can reduce it. *Curr Genet*, *68*(1), 39-48. https://doi.org/10.1007/s00294-021-01211-1

Solomon, W. C., Myint, W., Hou, S., Kanai, T., Tripathi, R., Kurt Yilmaz, N., Schiffer, C. A., & Matsuo, H. (2019). Mechanism for APOBEC3G catalytic exclusion of RNA and non-substrate DNA. *Nucleic Acids Res*, *47*(14), 7676-7689. https://doi.org/10.1093/nar/gkz550

St Martin, A., Salamango, D., Serebrenik, A., Shaban, N., Brown, W. L., Donati, F., Munagala, U., Conticello, S. G., & Harris, R. S. (2018). A fluorescent reporter for quantification and enrichment of DNA editing by APOBEC-Cas9 or cleavage by Cas9 in living cells. *Nucleic Acids Res*, *46*(14), e84. https://doi.org/10.1093/nar/gky332

Thuronyi, B. W., Koblan, L. W., Levy, J. M., Yeh, W. H., Zheng, C., Newby, G. A., Wilson, C., Bhaumik, M., Shubina-Oleinik, O., Holt, J. R., & Liu, D. R. (2019). Continuous evolution of base editors with expanded target compatibility and improved activity. *Nat Biotechnol*, *37*(9), 1070-1079. https://doi.org/10.1038/s41587-019-0193-0

Trevino, A. E., & Zhang, F. (2014). Genome editing using Cas9 nickases. *Methods Enzymol*, *546*, 161-174. https://doi.org/10.1016/B978-0-12-801185-0.00008-8

UNAIDS. (2024, 2 August). *Global HIV & AIDS statistics — Fact sheet*. Retrieved 17 February from https://www.unaids.org/en/resources/fact-sheet

Wang, W., Sun, L., Zhang, S., Zhang, H., Shi, J., Xu, T., & Li, K. (2017). Analysis and prediction of single-stranded and double-stranded DNA binding proteins based on protein sequences. *BMC Bioinformatics*, *18*(1), 300. https://doi.org/10.1186/s12859-017-1715-8

Wang, X., Li, J., Wang, Y., Yang, B., Wei, J., Wu, J., Wang, R., Huang, X., Chen, J., & Yang, L. (2018). Efficient base editing in methylated regions with a human APOBEC3A-Cas9 fusion. *Nat Biotechnol*, *36*(10), 946-949. https://doi.org/10.1038/nbt.4198

Wellhausen, N., O'Connell, R. P., Lesch, S., Engel, N. W., Rennels, A. K., Gonzales, D., Herbst, F., Young, R. M., Garcia, K. C., Weiner, D., June, C. H., & Gill, S. I. (2023). Epitope base editing CD45 in hematopoietic cells enables universal blood cancer immune therapy. *Sci Transl Med*, *15*(714), eadi1145. https://doi.org/10.1126/scitranslmed.adi1145

Wittkopp, C. J., Adolph, M. B., Wu, L. I., Chelico, L., & Emerman, M. (2016). A Single Nucleotide Polymorphism in Human APOBEC3C Enhances Restriction of Lentiviruses. *PLoS Pathog*, *12*(10), e1005865. https://doi.org/10.1371/journal.ppat.1005865

World Health Organization. (2024, 22 July). *HIV and AIDS*. Retrieved February 17 from https://www.who.int/news-room/fact-sheets/detail/hiv-aids

Zhang, C., Lu, Y. J., Wang, M., Chen, B., Xiong, F., Mitsopoulos, C., Rossanese, O., Li, X., & Clarke, P. A. (2024). Characterisation of APOBEC3B-Mediated RNA editing in breast cancer cells reveals regulatory roles of NEAT1 and MALAT1 lncRNAs. *Oncogene*, *43*(46), 3366-3377. https://doi.org/10.1038/s41388-024-03171-5

Zhang, C., Yang, Y., Qi, T., Zhang, Y., Hou, L., Wei, J., Yang, J., Shi, L., Ong, S. G., Wang, H., Wang, H., Yu, B., & Wang, Y. (2023). Prediction of base editor off-targets by deep learning. *Nat Commun*, *14*(1), 5358. https://doi.org/10.1038/s41467-023-41004-3

Zhang, H., Bamidele, N., Liu, P., Ojelabi, O., Gao, X. D., Rodriguez, T., Cheng, H., Kelly, K., Watts, J. K., Xie, J., Gao, G., Wolfe, S. A., Xue, W., & Sontheimer, E. J. (2022). Adenine Base Editing In Vivo with a Single Adeno-Associated Virus Vector. *GEN Biotechnol*, *1*(3), 285-299. https://doi.org/10.1089/genbio.2022.0015

Zhang, Y., Liu, Y., Qin, W., Zheng, S., Xiao, J., Xia, X., Yuan, X., Zeng, J., Shi, Y., Zhang, Y., Ma, H., Varshney, G. K., Fei, J. F., & Liu, Y. (2024). Cytosine base editors with increased PAM and deaminase motif flexibility for gene editing in zebrafish. *Nat Commun*, *15*(1), 9526. https://doi.org/10.1038/s41467-024-53735-y

Zhao, N., Zhou, J., Tao, T., Wang, Q., Tang, J., Li, D., Gou, S., Guan, Z., Olajide, J. S., Lin, J., Wang, S., Li, X., Zhou, J., Gao, Z., & Wang, G. (2024). Evolved cytidine and adenine base editors with high precision and minimized off-target activity by a continuous directed evolution system in mammalian cells. *Nat Commun*, *15*(1), 8140. https://doi.org/10.1038/s41467-024-52483-3

Zhou, L., Su, J., Long, J., Tao, R., Tang, W., Qin, F., Liu, N., Wang, Y., Jiao, Y., Hu, Y., Jiang, L., Li, L., Yang, Y., & Yao, S. (2022). A universal strategy for AAV delivery of base editors to correct genetic point mutations in neonatal PKU mice. *Mol Ther Methods Clin Dev*, *24*, 230-240. https://doi.org/10.1016/j.omtm.2022.01.001

Ziegler, S. J., Hu, Y., Devarkar, S. C., & Xiong, Y. (2019). APOBEC3A Loop 1 Is a Determinant for Single-Stranded DNA Binding and Deamination. *Biochemistry*, *58*(37), 3838-3847. https://doi.org/10.1021/acs.biochem.9b00394

Acknowledgments

First of all, I want to thank my orientator, Prof. Carsten Münk for the opportunity to realize this work at his laboratory and the highly competent advice I always received. I truly appreciate everything I learned. I would also like to thank my two co-orientators, Prof. Heiner Schaal and Prof. Ingo Drexler.

Special thanks also go to the AG Münk for supporting me during my time at the lab and all the important teachings. Thank you, Kannan Balakrishnan, for all your help, patience and advice, it was truly a pleasure to work with you. Anucha Sangwiman and Shenglin Zhu thank you for your help, the shared lunches and coffee breaks in between experiments. My gratitude for Lucía Cano Ortiz's support, patience and kindness goes beyond what I could have hoped for when I entered the lab. Aside from excellent scientific advice and fun Portuguese classes, I have learned more form you that I can ever thank you for. Lastly, I want to especially thank Wioletta Hörschken for all her help, excellt organization and encouragements over last few years.

I want to thank my loving parents, Barabara Tartarotti-Alfreider and Albin Alfreider, for supporting and encouraging me as always. Thank you for showing me from a very young age how interesting scientific research is. You taught me the importance of determination and courage to pursue my goals and dreams. I will be forever thankful for all that you have done for me.

My caring and loving grandmother, Ulrike Tartarotti, although already passed away, is a constant source of inspiration. I cannot express my eternal gratitude for my amazing grandfather, Siegfried Tartarotti. I am beyond thankful for all the opportunities you gave me through your help.

I want to thank all my friends I met in Düsseldorf at university, you all made the years spent in Düsseldorf into a truly specially time. I want to especially thank Luca Schuster and Jacqueline Nacken for all the encouragements, support, bike rides, and wonderful travels. Susanne Markl, my longest childhood friend, thank you for all your kindness and help as always.