International PhD Program in Biomolecular Sciences

Department of Cellular, Computational and Integrative Biology – CIBIO

XXXVI Cycle

Expanding the CRISPR-Cas9 toolbox for genome editing

Tutor

Prof. Anna Cereseto

University of Trento

CIBIO department

Advisor

Antonio Casini, PhD

Alia Therapeutics

Ph.D. Thesis of

Eleonora Pedrazzoli

University of Trento

CIBIO department

Academic Year 2022-2023

Declaration of Authorship

I, Eleonora Pedrazzoli, confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Contribution to the experiments

Bioinformatic analysis

All the bioinformatic pipelines and downstream analyses were developed and/or performed by Matteo Ciciani (CIBIO, University of Trento).

Part I - Development of rCas9HF, an optimized SpCas9 high-fidelity variant

Andrea Bianchi and I produced and purified the Cas9 proteins used in all the experiments. Andrea Bianchi (CIBIO, University of Trento) conducted the experiments prior to the selection of rCas9HF, helped by Kalina Aleksandra Badowska (Alia Therapeutics, Trento) for genome editing analyses. I performed the RNP-based electroporation experiments to compare rCas9HF, HiFiCas9 and WT Cas9 in U2OS cells, the deep-sequencing editing analyses and GUIDE-seq. Giacomo Frati and Mégane Brusson (IMAGINE Institute, Paris) performed the electroporation of primary cells from sickle cell disease patients, I analyzed the editing efficiency of these samples. Alessandro Umbach and Simone Amistadi (CIBIO, University of Trento) performed HDR experiments and helped me with the NGS libraries preparation. Daniele Arosio (Biophysics Institute, National Research Council of Italy, Trento) analyzed the protein structure.

Part II - CoCas9, a compact nuclease from the human microbiome for efficient and precise genome editing

Michele Demozzi, Elisabetta Visentin, Ilaria Bonuzzi (CIBIO, University of Trento) and I performed *in vitro* PAM detection and EGFP-disruption assays. I analyzed the ends generated by the CoCas9 cleavage, designed and compared the CoCas9 sgRNA scaffold versions. Elisabetta Visentin performed the test for CoCas9 spacer optimization. Elisabetta Visentin and I performed CoCas9 and SpCas9 nuclease activity comparisons. Ilaria Bonuzzi performed the western blots. I performed the GUIDE-seq for CoCas9 and SpCas9. Elisabetta Visentin, Simone Amistadi, and Lorenzo Lucchetta (CIBIO, University of Trento) performed the experiments to compare CoCas9 with small Cas9 orthologs. Giulia Maule (CIBIO, University of Trento) contributed to GUIDE-seq analyses. I designed and tested CoABE8e and compared it with other base editors, Ilaria Bonuzzi helped me in the base

editing efficiency experiments and analyses. Elisabetta Visentin, Ilaria Bonuzzi and I performed the experiments on primary cells. Laura Pezzè (Alia Therapeutics, Trento) helped me in the AAV vectors design, production, and testing. Federica Esposito (TIGEM, Napoli) performed the *in vivo* experiments, while I analyzed the editing efficiency on these samples.

Elisabetta Visentin, Ilaria Bonuzzi, and Lorenzo Lucchetta contributed to this work as part of their master thesis training period, during which I served as their tutor.

The first part of this thesis work has been published in
Pedrazzoli, E., Bianchi A. et al. An optimized SpCas9 high-fidelity variant for direct protein delivery. Mol Ther 31, 2257–2265 (2023)
where I am listed as first co-author[1] and contributed to the writing.

The second part has been published in
Pedrazzoli, E., Demozzi, M., Visentin, E. et al. CoCas9 is a compact nuclease from the human microbiome for efficient and precise genome editing. Nat Commun 15, 3478 (2024)
where I am listed as first co-author[2] and contributed to the writing.
The second part is comprised in the patent
"TYPE II CAS PROTEINS AND APPLICATIONS THEREOF", International Application No. PCT/EP2022/087314, Publication Date 29.06.2023. Inventors: CASINI, Antonio; CERESETO, Anna; SEGATA, Nicola; DEMOZZI, Michele; PEDRAZZOLI, Eleonora; CICIANI, Matteo; VISENTIN, Elisabetta; PEZZÈ, Laura.

Trento, 26th March 2024

Abstract

The transformative impact of CRISPR-Cas systems has revolutionized the landscape of genome editing research, making unprecedented strides into clinical settings for treating diseases for which there are no treatments.

Nonetheless, the CRISPR toolbox still needs to be advanced to tackle the heterogeneity of genetic diseases and the obstacles of *in vivo* therapeutic applications. Primary problems when using CRISPR systems concern not only the balance between efficiency and precision, but also size, which becomes critical when targeting specific organs depending on adeno-associated viral (AAV) vectors for delivery.

This thesis work aimed to enlarge the CRISPR-based editing toolbox by introducing new Cas9 proteins. Two approaches were exploited: 1) rational molecular engineering to obtain a high-fidelity version of the *Streptococcus pyogenes* Cas9 (SpCas9), and 2) the identification of new CRISPR-Cas9 loci from metagenomic data.

In the first approach, a high-fidelity SpCas9 variant (rCas9HF) was generated starting from the pool of amino acid substitutions resulting from the screening of evoCas9. Differently from evoCas9, rCas9HF is characterized by the K526D substitution, which allows efficient editing through ribonucleoprotein (RNP) electroporation. rCas9HF indels generation and off-target effects are comparable to HiFi Cas9, which is currently the only available high-fidelity Cas9 that can be used as RNP. The two high-fidelity mutants were also tested for homology-directed repair (HDR) and ability to modify the genome in CD34+ cells, showing that the variants assure a benefit compared to wild-type (wt) SpCas9 in terms of precision. Moreover, rCas9HF and HiFi Cas9 are characterized by different editing profiles, providing a valuable new precise tool for genome editing.

In the second approach, CoCas9 was identified from a massively expanded microbiome repository as a nuclease characterized by compact molecular size coupled with an editing efficiency on genomic loci slightly similar to SpCas9. CoCas9 was demonstrated to be very precise, compatible with base editing technology, and deliverable *in vivo* through a single AAV, as demonstrated in transducing experiments targeting the mouse retina. When compared to other <1100 amino acid (aa) Cas9s, this new ortholog showed better or comparable on-target efficiency, with a better specificity profile and broader PAM.

Overall, this thesis uncovers and characterizes two Cas9s, further enriching the CRISPR toolbox, and thus improving the ability to efficiently target the variety of pathogenic mutations in the human genome.

Index

List of acronyms and abbreviations....8
Introduction....9
Genome editing B.C. (before CRISPR)....9
DNA repair mechanisms....9
Meganucleases, ZFNs and TALENs....11
The revolution of the CRISPR system....13
History of the CRISPR discovery....13
CRISPR-Cas system biology....14
The power of engineered CRISPR systems....20
Overcoming the SpCas9 limitations....21
Avoiding DSBs....21
High-fidelity Cas proteins generation....24
Delivery of CRISPR-Cas systems....25
Expanding the targetable genome fraction: modify the PAM....26
Developing and finding new CRISPR-Cas9 systems....27
Aim of the thesis....29
Part I - Development of rCas9HF, an optimized SpCas9 high-fidelity variant....30
Results....30
Part II - CoCas9, a compact nuclease from the human microbiome for efficient and precise genome editing....50
Methods....51
Results....59
Identification of CRISPR-Cas9 systems from bacterial metagenomes....59
Characterization of the CoCas9-based tools....61
CRISPR-CoCas9 in relevant cellular and animal models....72
Discussion....74
Bibliography....78
Appendix....89

List of acronyms and abbreviations

aa	amino acid
AAV	Adeno-Associated Virus
bp	base pair
Cas	CRISPR-ASsociated (proteins)
CBE/ABE	Cytosine/Adenine Base Editor
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
crRNA	CRISPR RNA
dCas9	dead Cas9
DSB	Double Strand Break
GUIDE-seq	Genome-wide, Unbiased Identification of DSBs Enabled by Sequencing
HDR	Homology-Directed Repair
Indels	Insertion-Deletions
KO	Knock-Out
LNP	Lipid NanoParticles
nCas9	nickase Cas9
NGS	Next Generation Sequencing
NHEJ	Non-Homologous End Joining
NLS	Nuclear Localization Signal
PAM	Protospacer-adjacent Motif
PE	Prime Editor
PI	PAM-Interacting (domain)
rCas9HF	recombinant Cas9 High-Fidelity
RNP	RiboNucleoProtein
sgRNA	single-guide RNA
TadA	tRNA adenosine deaminase
TALEN	Transcription Activator-Like Effector Nuclease
TIDE	Tracking of Indels by Decomposition
tracrRNA	trans-activating CRISPR RNA
ZFN	Zinc Finger Nuclease
wt	wild-type

Introduction

Genome editing B.C. (before CRISPR)

In 2023, the 70th anniversary of the discovery of DNA structure is commemorated and it is possible to trace to this breakthrough the very beginning of the genome engineering field, which arose from the curiosity and necessity of analyzing the consequences of DNA sequence modifications[3]. Genome editing is defined as the modification of the genetic material of living cells through cutting, inserting, or otherwise altering DNA[4]. Among others, two are the essential components: a molecular machinery that has access to the DNA and the ability of cells to repair DNA damage. The beginning of the genome editing journey started when researchers understood how to disrupt the DNA and learned how to analyze the variations in the genomic sequence after the DNA repair occurred.

DNA repair mechanisms

Editing the genome needs the DNA to be accessible. In 1994, Maria Jasin showed how a cut in the DNA of mammalian cells performed with the meganuclease I-*Sce*I was effectively repaired through non-homologous end joining (NHEJ) or HDR[5]. This was not only the first indication that the genomic sequence could be modified using exogenous DNA, but also that the double-strand breaks (DSB) in the genome sensibly improve the rate of HDR[5,6].

DSBs naturally occur in the genome as a consequence of cellular processes, such as meiosis and DNA replication, but most result from damaging causes, which include reactive oxygen species, ionizing radiation, and inadvertent action of nuclear enzymes[7,8]. In DSBs, the phosphate backbones of the two complementary DNA strands are broken simultaneously, making DSB one of the most cytotoxic lesions for the DNA. Depending on the cause of the DSB and the cell cycle phase in which it occurs, NHEJ and HDR are the two main mechanisms of DSB repair.

The predominant form of repair outside of the S and G2 phases is NHEJ, which only needs 4 bp of microhomology for the repair to bind incompatible ends through random insertions and deletion (indels) at the cleavage site[9] (**Fig. 1**). Ku proteins form a heterodimer that covers the DNA ends, then the DNA-dependent protein kinase catalytic subunit (DNA-PKcs) binds the Ku proteins. At this point, Artemis is involved, as the most used endonuclease during the NHEJ process. After the action of Artemis, nucleotide addition occurs by the Pol X family polymerases; finally, the DNA ligase 4 complex carries out the critical ligation step[7,10]. The participation of multiple enzymes makes NHEJ an extremely flexible and rapid system to repair DNA without compromising its integrity.

On the other hand, HDR exploits the redundancy of genetic information in the form of sister chromatids or homologous chromosomes to fix compromised DNA[11] (**Fig. 1**). To have availability of homologous sequences, HDR occurs during the S/G2 cell cycle phases, and it is characterized by DNA resection, mainly driven by the MRN complex. MRN 3′−5′ exonuclease activity displaces Ku proteins from the DNA end and the association of RAD51 to the single-strand filament is allowed. The RAD51 nucleoprotein filament invades a homologous sequence, with subsequent strand extension[12]. In gene editing, HDR-mediated repair can be used to correct or substitute a mutated gene providing a donor containing the correct copy of the whole gene or part of it. The donor must be flanked by homology arms to allow HDR, which restores the mutated gene by replacing it with the correct sequence[13].

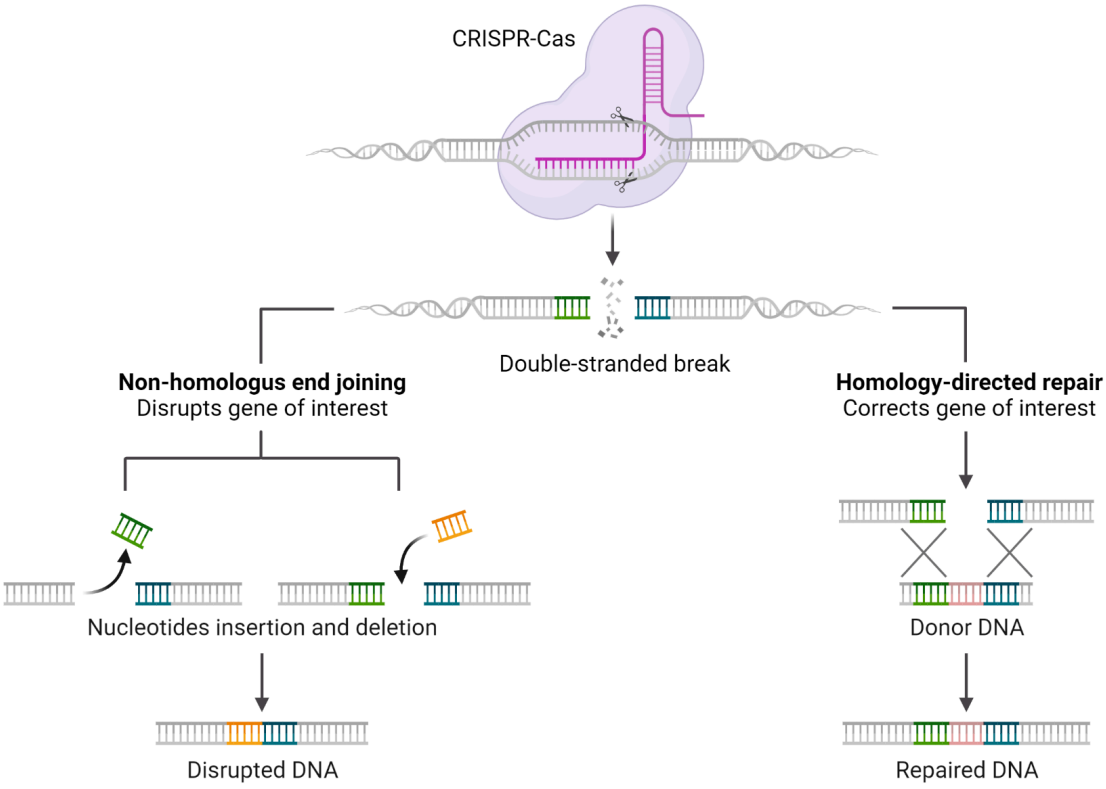

Figure 1. Schematic representation of NHEJ and HDR. *Adapted from "CRISPR-Cas9 Gene Editing", by BioRender.com (2023).*

In consideration of their fundamental attributes, NHEJ and HDR exhibit distinctive advantages and drawbacks. NHEJ is an efficient repair pathway that operates throughout the cell cycle, making it particularly advantageous in non-dividing cells or during the G0 phase. However, its error-prone nature may introduce small indels at the repair site, potentially leading to gene disruptions. In contrast, HDR is a high-fidelity repair pathway that relies on homologous sequences for precise template-directed repair. This makes HDR ideal

for maintaining genomic integrity, but its dependency on the cell's homologous recombination machinery restricts its activity to the S and G2 phases, limiting its applicability in non-dividing or slowly dividing cells. Additionally, HDR is less efficient than NHEJ, and the introduction of exogenous repair templates can pose challenges. The choice between NHEJ and HDR often depends on the specific cellular context and the desired outcome, reflecting a trade-off between repair speed, accuracy, and applicability across various cell types and stages of the cell cycle.

Meganucleases, ZFNs and TALENs

The cell utilizes repair mechanisms to fix damaged DNA, and targeted DSBs in specific regions allow for modifications of the genomic sequence depending on the type of repair that occurs. Consequently, engineered nucleases capable of creating breaks in the DNA are ideal tools for genome editing. Before the discovery of CRISPR-Cas9, meganucleases, zinc finger nucleases (ZFN), and transcription activator-like effector (TALE)-nucleases (TALEN) played a vital role as programmable endonucleases (**Fig. 2**).

Meganucleases naturally perform recognition and self-splicing activities, particularly well-characterized in the LAGLIDADG family[14] (**Fig. 2a**). Despite their appealing attributes, such as low toxicity[15,16], meganucleases are considered low-frequency cutters because the recognized sequence is typically 14-30 bp[17], making these endonucleases both highly specific and inflexible for genome editing. After the year 2000, efforts were made to engineer meganucleases for recognition of different genomic sequences[18,19], but reconfiguring the protein structure for each target proved to be a demanding process that may not yield commensurate benefits.

Differently from meganucleases, ZFNs and TALENs are synthetic proteins. ZFNs are chimeric proteins designed to specifically target and modify DNA sequences in a highly precise manner. Published in 1996, ZFNs are composed of a zinc-finger domains array and a non-specific FokI catalytic domain fused at the carboxy terminus[20]. ZFNs act as subunits: the FokI catalytic domain cleaves the DNA once dimerized, therefore two ZFNs recognizing the DNA sequence upstream and downstream of the cut are needed (**Fig. 2b**)[21]. The ZF domain comprehends 3-6 fingers[22], so that a ZFN system is able to specifically target a DNA sequence up to 36 bp, giving this technology remarkable accuracy[6]. Even if ZFNs are demonstrated to be a promising genome editing tool thanks to the possibility of reprogramming the targeting domain, the recognition of three base pairs (bp) per module remains an open problem for the use of these chimeric proteins, since the pairing between ZF and DNA is not perfectly modular[23]. Moreover, the rearrangement of ZFNs is a notably

time-consuming and resource-intensive process, causing the limited use of the technology across the laboratories.

In order to simplify the interaction between the genomic sequence and the DNA binding domain of ZFNs, some groups came up with the idea of fusing the FokI nuclease to an array of transcription activator-like effectors (TALEs)[24–27]. Conceptually similar to ZF domains, TALE proteins can be used to bind a specific DNA sequence, with the significant difference of one TALE being able to recognize a single nucleotide. The DNA binding domain contains a repeated highly conserved 33-34 amino acid sequence except for the 12th and 13th aa. These two locations are highly variable (Repeat Variable Di-residue or RVDs) and show a strong correlation with specific nucleotide recognition[28]. Four RVD domains recognize guanine, adenine, cytosine, and thymine, thus each TALE module recognizes a single DNA base[29]. Due to FokI catalytic domain fused to the C-terminus of TALE, TALENs work as pairs as well, and the TALE arrays must be designed on opposite strands (**Fig. 2c**). This characteristic ease of generation makes TALENs a more accessible and practical choice in comparison to ZFNs, where the task of crafting a distinct protein for every target remains a notable obstacle[30].

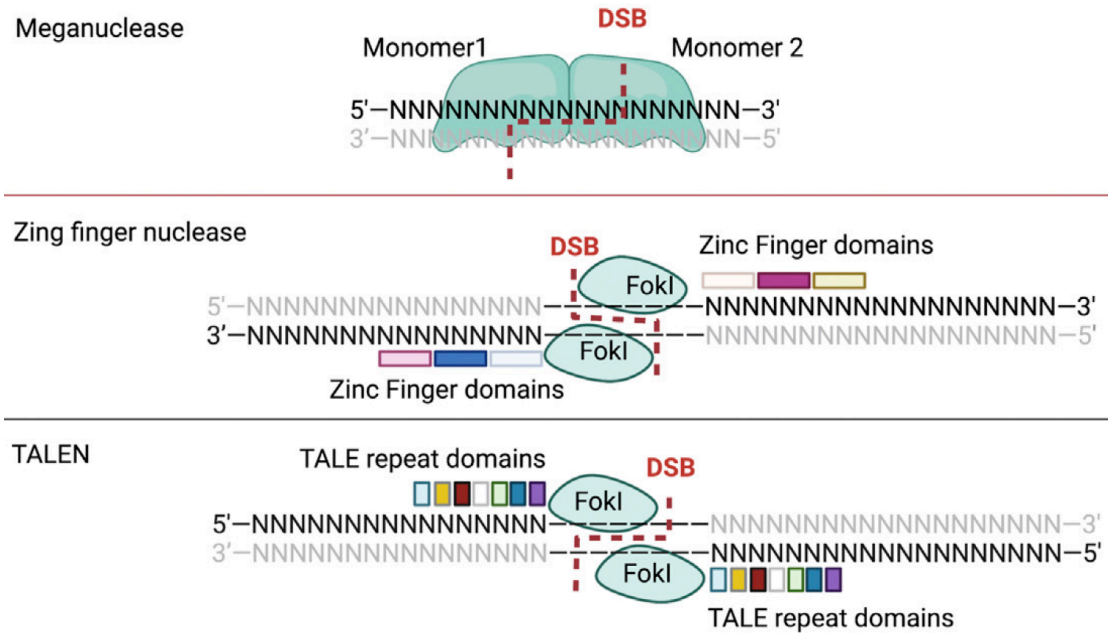

Figure 2. Simplified representation of meganucleases, ZFNs, and TALENs. *Adapted from Lu et al., 2022[31].*

The revolution of the CRISPR system

History of the CRISPR discovery

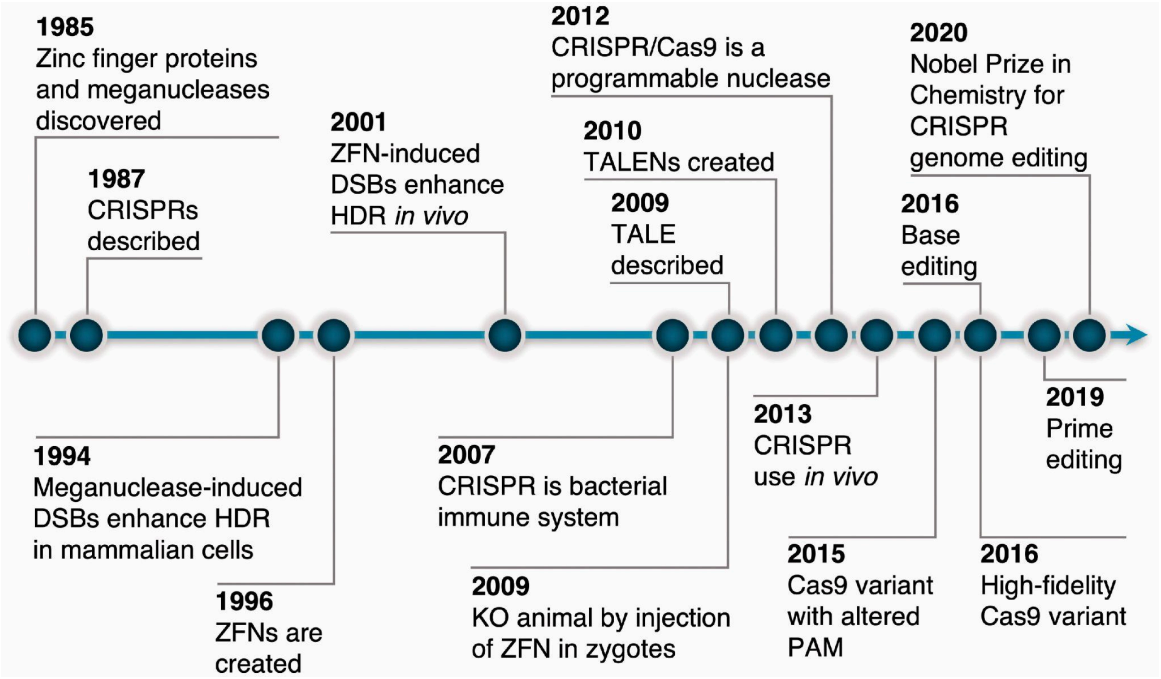

Figure 3. Key developments in the history of genome editing. *From Troder and Zevnik, 2022*[15].

The CRISPR-Cas system emerged in 2012, in a scientific scenario where genome editing was achievable but required significant resources and was limited to specific laboratories, although its story began in the 80s (**Fig. 3**).

In 1987, an unusual repetitive DNA sequence was identified in the *Escherichia coli* genome by Ishino *et al.*[32]. Few years later, Francisco Mojica observed a similar pattern of repeated sequences, this time within archaea[33]. The presence of these recurring sequences in vastly distinct organisms persuaded Mojica they held biological significance. At the beginning of the new millennium, the "genomics era" emerged, with the possibility of sequencing batches of genomes concurrently and consulting an increasing number of sequencing databases. In this context, unique sequences were noticed alternating the repeats in the genomes of *E. coli*, *Streptococcus thermophilus,* and *Yersinia pestis*[34–36]. These sequences were called "spacers" and matched particular viral sequences, a critical characteristic in the intuition of CRISPR being part of the prokaryotic adaptive system[37]. In those years, Bolotin *et al.* noted that the spacers shared a common sequence at one end; this sequence, the protospacer adjacent motif (PAM), was discovered to be required for target recognition[35].

In an experimental study conducted by Horvath and colleagues at Danisco, it was demonstrated that CRISPR systems indeed function as an adaptive immune system in *S. thermophilus*. They observed the integration of new phage DNA into the CRISPR array, enabling the bacterium to defend against subsequent phage attacks, as elucidated by Barrangou *et al*. in 2007[38]. Moreover, their research indicated that Cas9 was likely the sole protein necessary for interference, the mechanism by which the CRISPR system neutralizes invading phages, although specific details of this process were not yet fully understood[38].

In 2008, John Van Der Oost laboratory contributed to the CRISPR studies by describing how spacer sequences are transcribed into CRISPR RNAs (crRNA), that guide Cas proteins to the target DNA[39]. To fully figure out the role of RNA in the CRISPR system, Emmanuelle Charpentier and colleagues demonstrated a second small RNA, the trans-activating CRISPR RNA (tracrRNA)[40], which forms a duplex with the crRNA, and this RNA complex guides the Cas9 to its target.

Once the main characters of the CRISPR-based adaptive immunity were clarified, the next step was understanding if the system would be useful for editing DNA in an extra-bacterial environment. In 2012, two similar papers came out: in June the work from Doudna and Charpentier on Science[41], and in September the one from Siksnys on PNAS[42]. Both of them focused on the biochemical characterization of Cas9-mediated cleavage. They verified the cleavage site and the requirement for the PAM and showed how the Cas9 domains perform the DNA cut. Remarkably, they demonstrated that Cas9 DSBs can be easily reprogrammed by changing the crRNA sequence[41,42]. In Jinek *et al*. they engineered the dual crRNA-tracrRNA guide in order to generate a synthetic single guide RNA (sgRNA) to simplify the use of the system for genome editing. After this incredible breakthrough, the “CRISPR era” officially began after the publication of various articles where it was shown the successful editing with CRISPR-Cas9 in eukaryotic cells[43–45].

CRISPR-Cas system biology

Over the past 36 years since the initial discovery of CRISPR, our understanding of the CRISPR-Cas system has significantly increased. This system is part of various pathways developed by bacteria and archaea to restrict incoming exogenous nucleic acids. These organisms are constantly exposed to unknown genetic material, which can offer positive features, such as antibiotic resistance genes encoded by temperate phages, or lead to fitness compromise, as seen in virulent bacteriophage infections[46]. Fundamentally, CRISPR-Cas is an RNA-based, adaptive, and heritable system, crucial for both the protection and evolution of prokaryotes[47], and it operates through the cooperation of two elements: the CRISPR array and *cas* (CRISPR-associated proteins) genes[48].

The CRISPR locus and systems classification. A CRISPR array is a highly variable region composed of 30-40 bp non-contiguous repeats, depending on the prokaryotic species, alternating variable sequences of similar size, the spacers[49,50]. Spacers match sequences in bacteriophage genomic material or other types of exogenous DNA[35], while the repeat serves as a scaffold in the final gRNA structure[51]. At the 5' end of the repeat-spacer array, the leader sequence contains the promoter necessary for transcription of the array, and it is crucial for the acquisition of new spacers[52,53]. The *cas* genes provide the enzymatic tools for all the immunity steps and are organized in operons, usually found adjacent to the CRISPR[54]. As shown in **Figure 4a**, a typical *cas* genes set comprises *cas1*, *cas2,* and a specific combination of core and subtype specific Cas proteins[55].

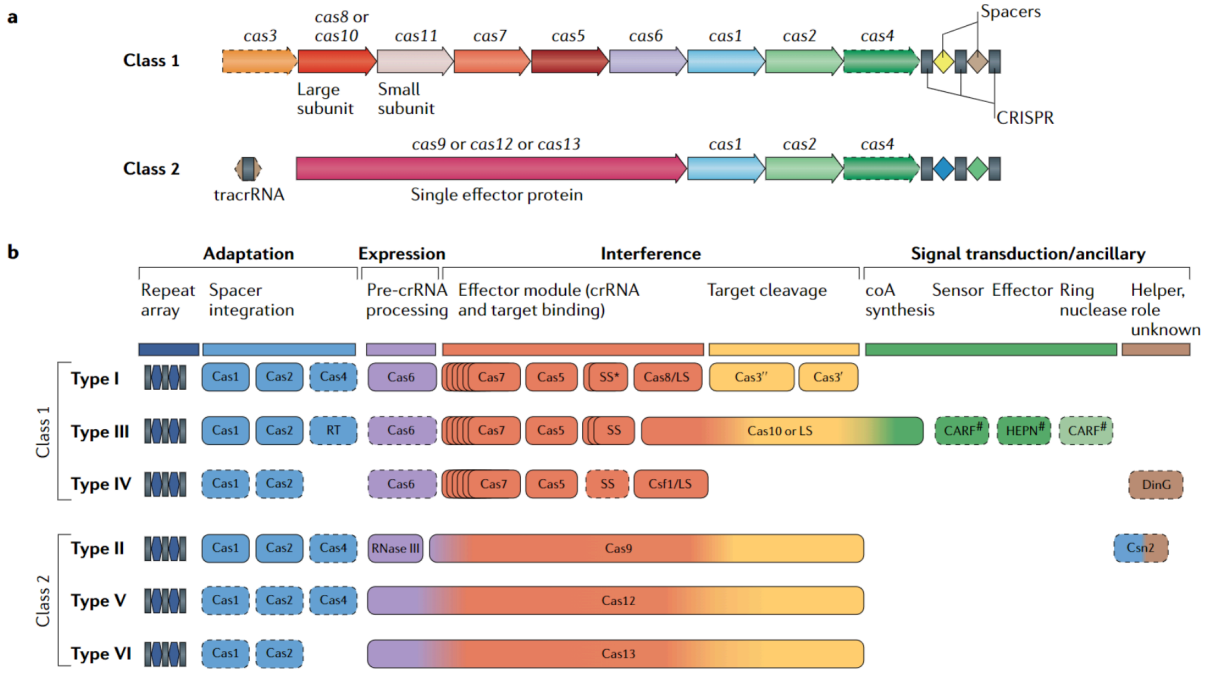

Figure 4. The modular organization of CRISPR–Cas systems into two classes and six types. *(**a**) Standard configurations of class 1 and class 2 CRISPR loci. (**b**) Structural units of the six CRISPR-Cas types divided by their functions in the CRISPR-based adaptive immunity (adaption, expression, interference, and signal transduction/ancillary). Dashed outlines indicate dispensable and/or missing components. From Makarova et al., 2020[56].*

Being a crucial component of prokaryotic immunity, CRISPR-Cas systems exhibit remarkable variability and rapid evolution, making the classification process challenging and pertinent for only brief intervals. Given these challenges, the classification described in Makarova *et al*. in 2020 relies on four primary criteria: the presence of signature Cas proteins, the degree of sequence homology among Cas proteins, the evolutionary

relationships among the most conserved Cas proteins, and the preservation of locus organization[56]. In their work, two CRISPR classes have been categorized into six types (**Fig. 4b**), with over 33 subtypes distinguished by variations in their interference module composition and their specificity in target nucleic acids.

The *cas* genes are categorized into three functional modules[55]: the adaptation module, which includes Cas1 integrase, Cas2, and sometimes Cas4; the expression processing module, responsible for pre-crRNA processing, which can be carried out by Cas6, bacterial RNase III, or effector Cas proteins; and the interference or effector module, which plays a role in target recognition and nucleic acid cleavage.

These modules, though distinct, can overlap in their functions. The macroscopic difference between the two classes is that Class 1 is characterized by a multisubunit effector complex called Cascade (e.g. composed of Cas6, Cas7, and Cas5), while class 2 effector module consists of a single protein (e.g. Cas9, Cas12, and Cas13). In both classes, the effector module forms a crRNA-binding complex to bind and process the target.

The 90% of CRISPR loci found in bacteria and archaea belong to Class 1, which is divided into type I, type III, and type IV[57]. All type I systems are characterized by the presence of the *cas3* gene. Cas3 is an ATP-dependent translocase/helicase enzyme with the role of unwinding DNA and RNA-DNA strands, which facilitates DNA target cleavage[58]. On the other hand, type III systems can recognize both DNA and RNA, and Cas10 is the signature protein, with an RNA recognition motif called Palm and a cyclase domain responsible for cutting[59]. Lastly, in type IV the effector module is much smaller and does not contain domains capable of target cleavage or spacer insertion[60]. Even if it was demonstrated that it is possible to edit human cells using Class 1 CRISPR system[57], it remains largely unexplored for genome engineering applications because of the need for the Cascade complex.

The Class 2 CRISPR-Cas system finds extensive application in both basic and translational biomedical applications, thanks to the presence of a single effector nuclease. Cas12 is the preferred endonuclease in type V systems: just like type II systems, Cas12 can target only DNA and relies on a tracrRNA to carry out its functions[61]. In contrast with the other types, type VI focuses on editing RNA, with Cas13 serving as the pivotal endonuclease for RNA editing[61]. The most characterized genome editing tool, SpCas9, is part of Class 2 type II. The Cas9 protein characterizes all type II CRISPR-Cas systems and it is involved in the adaptation phase, in crRNA processing and target DNA cleavage assisted by crRNA and tracrRNA[62–64]. Type II CRISPR-Cas systems use two extra-CRISPR elements for the pre-crRNA processing, RNase III and tracrRNA. Similarly to type V and VI, type II systems discern between self and non-self DNA by identification of the PAM element that is not present in the CRISPR locus, then no spacers could be a target for destruction[65]. Type II

systems are classified into 3 subtypes: type II-A systems include Csn2, type II-Bs are identified by Cas4, while type II-Cs are typically recognized for the absence of both Cas4 and Csn2[66]. The functioning of subtype II-A is the most characterized, being the group SpCas9 belongs to. Together with Cas1, Cas2, and Cas9, type II-A includes the subtype signature protein Csn2, considered a subtype-specific signature Cas protein, involved in the adaptation stage[67,68]. The specific gene of the cas effector complex in type II-B is *cas4*, which is also found in subtypes I-A to I-D Cas4. Cas4 proteins exhibit 5′-single-stranded DNA exonuclease activity and were demonstrated to assist the Cas1-Cas2 complex in the selection of new spacers with PAM-based interference in type I-D[69]. Finally, type II-C CRISPR-Cas systems have only three *cas* genes (*cas1*, *cas2*, and *cas9*)[66].

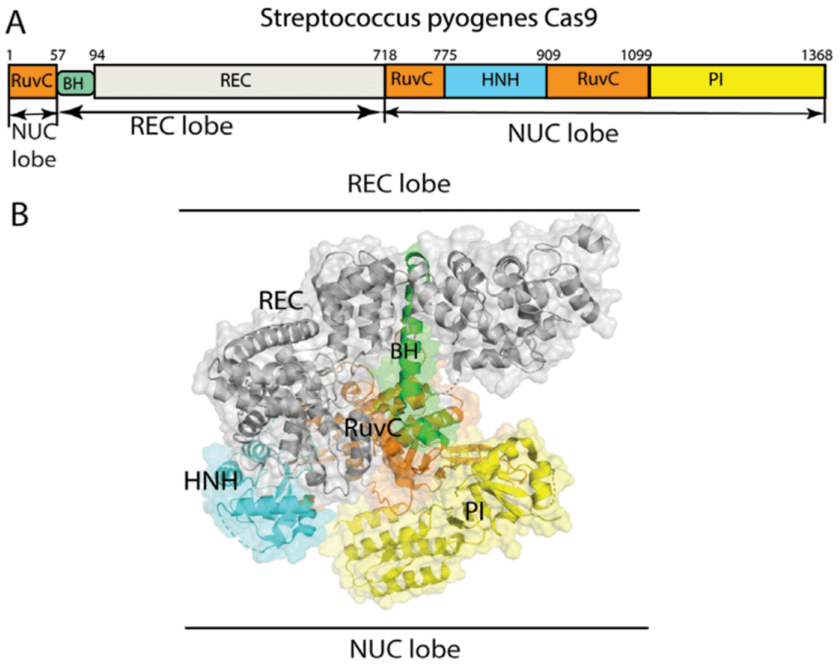

Figure 5. Structure of SpCas9. *Organization of type II SpCas9 domains and cartoon-surface representations of the three-dimensional structure of SpCas9. From Leonova et al., 2020[70].*

SpCas9 structure. The most characterized Cas9 is SpCas9 whose structure was defined by Nishimasu and colleagues[71], providing insights on the molecular mechanism of this enzyme. The SpCas9 general structure presents two lobes, one for the nuclease activity (NUC), and the other one for the DNA recognition (REC). Within the NUC lobe, two nuclease domains reside: RuvC and HNH, which cut the DNA strand non-complementary to the gRNA and the complementary strand, respectively[56]. Additionally, in the nuclease lobe the variable C-terminal domain (CTD) is present, responsible for PAM recognition and binding, known as

the PAM-interacting (PI) domain (**Fig. 5**). The REC lobe is responsible for DNA complexing and activation of the HNH catalytic site of Cas9 after binding to the target[72].

The CRISPR-based adaptive immunity system. The adaptive immunity mediated by CRISPR-Cas systems involves three common steps: adaptation of CRISPRs, expression, and interference[73] (**Fig. 6**). For the scope of this thesis, the description will focus especially on type II CRISPR systems.

During the adaptation, a foreign sequence, called “protospacer”, is recognized by the Cas system thanks to the presence of the PAM, typical for each immune system[74]. The protospacer is captured and integrated as a spacer in a CRISPR array thanks to the conserved protein machinery Cas1-Cas2[67]. In type II systems, Cas9 is involved in PAM recognition and protospacer selection[62]. The PAM within the pre-spacer ensures integration into the CRISPR in the correct orientation, allowing target recognition during the next invasion[75]. Cas1-mediated processing of the pre-spacer creates two 3’-OH ends required for nucleophilic attack on each strand of the leader-proximal repeat. The initial nucleophilic attack most likely occurs at the leader-repeat junction and then a second attack at the existing repeat-spacer junction generates the full-site integration product and ensures accurate repeat duplication[76]. New spacers are directionally included in the CRISPR array starting from the region adjacent to the leader. This manner of spacer addition produces a “chronological memory” of the encounters between phages and hosts.

In the second phase, transcription of CRISPR from the leader region generates the crRNA, which is responsible for the specificity of CRISPR-Cas immunity. In particular, a long pre-crRNA is transcribed and processed into 30-65 bp crRNAs by specific endoribonucleases (e.g. Cas6); these mature crRNAs are composed by a spacer adjacent to a repeat. crRNAs associate with one or more Cas proteins to form effector complexes that target exogenous nucleic acids through crRNA:target sequence-specific recognition[77]. Type II CRISPR-Cas system has a peculiar pre-crRNA processing, which requires the concerted action of a small tracrRNA, the protein Cas9, and the RNA endoribonuclease III (RNase III) acting effector. tracrRNA is recognized and bound by Cas9 and simultaneously tracrRNA hybridizes the pre-crRNA. Secondly, RNase III identifies the formation of this dsRNA that is cleaved at one end. Once processed, crRNA remains associated with Cas9 thanks to the tracrRNA[40].

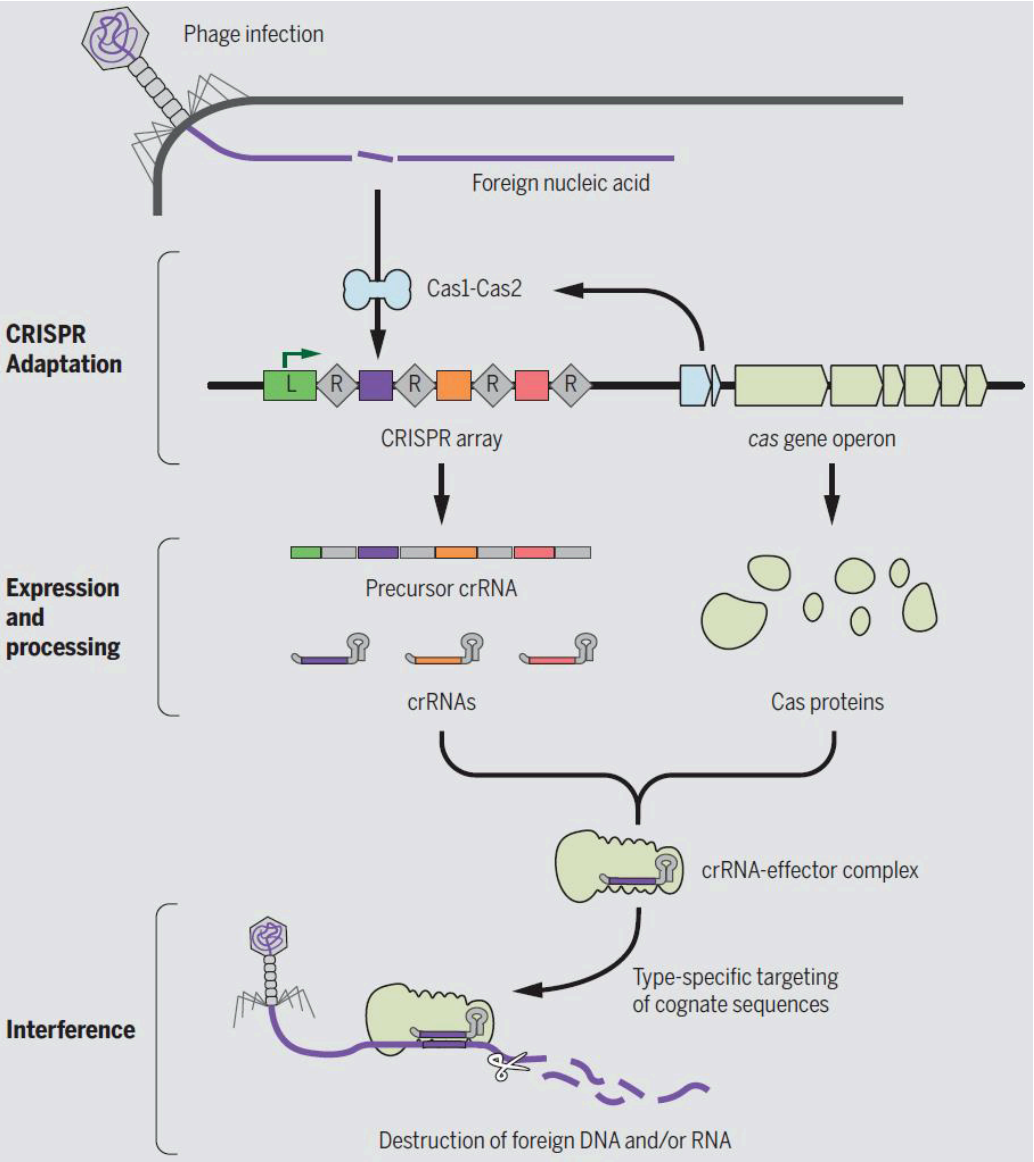

Figure 6. Overview of CRISPR-Cas-mediated adaptive immunity. *After phage infection, the first stage of CRISPR-Cas defense is CRISPR adaptation. This involves the incorporation of small fragments of foreign nucleic acid into the CRISPR array, stored as spacers (colored squares) between repeat sequences (R). The process is catalyzed by the Cas1-Cas2 complex. The second stage involves transcription of the CRISPR array and subsequent processing of the precursor transcript to generate crRNAs. Each crRNA contains a single spacer unit typically flanked by parts of the adjoining repeat sequence. In the interference phase, individual crRNAs assemble with Cas effector proteins to form crRNA-effector complexes. The crRNA-effector complexes catalyze the sequence-specific recognition and destruction of foreign DNA and/or RNA elements. From Jackson et al., 2017*[76].

The target interference step requires the identification of non-self sequences that are complementary to spacers in CRISPR loci. Generally, protospacer sequences flanked by a PAM sequence are recognized by a CRISPR ribonucleoprotein complex, consisting of the effector protein and crRNA guide: the recognition of the target is based on the annealing between the “seed” (7-13 bp sequence in proximity to the PAM) on the crRNA guide and the exogenous DNA[78]. Type II systems require only Cas9 bound to the crRNA guide base-paired with the tracrRNA for interference. The formation of this RNP complex triggers a conformational change of HNH and RuvC nuclease domains of Cas9, which adopt an active state that allows the interference by cleavage of both target strands, typically located 3 bp from the 3’ end of the protospacer[79].

The power of engineered CRISPR systems

The CRISPR-Cas9 system requires only three components for site-specific DNA interference: the Cas9 protein, the crRNA guide, and the tracrRNA (**Fig. 7a**). This characteristic and its adaptability made it the best choice for developing programmable genome editing tools. In Le Cong *et al*. they describe the codon-optimization of SpCas9 and the addition of nuclear localization signals (NLS) to ensure nuclear compartmentalization[43]. To reconstitute the non-coding RNA components, they expressed a tracrRNA and a pre-crRNA, comprising a single guide spacer flanked by direct repeats, under two different RNA polymerase III U6 promoters. Their initial spacer targeted a 30 bp “protospacer” in the human *EMX1* locus that precedes a 5’-NGG PAM. In two studies, Mali and Jinek demonstrated the possibility of editing human cells by synthesizing a human codon–optimized version of the Cas9 protein bearing nuclear localization signals and cloning it into a mammalian expression system. To direct Cas9 to cleave sequences of interest, they expressed crRNA-tracrRNA fusion transcripts, the single-guide RNA (sgRNA), under the control of the human U6 polymerase III promoter (**Fig. 7b**). This approach is constrained only by U6 transcription, which requires an initial G and the PAM sequence 5’-NGG following the 20 bp sgRNA target[44,45]. The sgRNA is composed of a seed sequence, at the 3’ end of the guide, and a non-seed sequence; the seed sequence and the PAM determine Cas9 specificity. In general, mismatches of one to five bp at the 5’ end of sgRNAs are better tolerated than those at the 3’ end. Single and double mismatches are tolerated to various degrees depending on their position along the guide RNA-DNA interface[80]. In CRISPR immunity, tolerating multiple mismatches between the 5’ end of crRNA and target DNA is an advantage because they allow spacers gained from one virus to be used in DNA interference to related viruses that carry closely related protospacers[81]. Starting from this first engineered CRISPR system, the use of the tool expanded both in terms of application and

diversification, being Cas9 the progenitor of a family of CRISPR-based genome editing technologies.

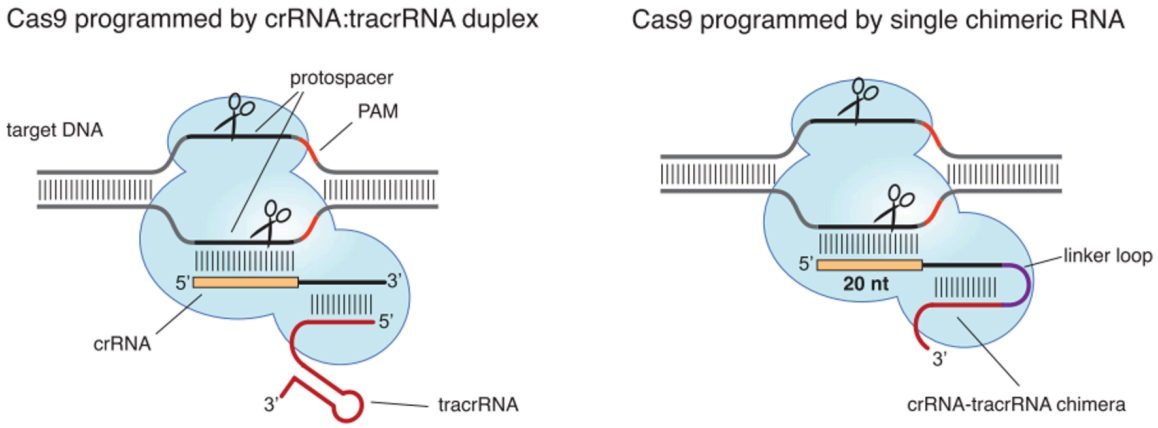

Figure 7. Type II wt CRISPR-Cas9 and reprogrammed system. *(**a**) In type II CRISPR-Cas systems, crRNA and tracrRNA guide Cas9 to the DNA targeting. (**b**) A chimeric RNA generated by fusing the 3' end of crRNA to the 5' end of tracrRNA guides the Cas9 cleavage in an engineered CRISPR-Cas9 system. From Jinek et al., 2012*[41].

Overcoming the SpCas9 limitations

The breadth of disorders necessitates tailored tools to address them effectively. Even though SpCas9 is the most used Cas in genome editing applications, its major limitations must be considered: the generation of DSBs, the specificity, the difficulty in using common viral vectors for delivery caused by SpCas9 size (1368 aa, 4.10 Kb), and the targetable sites.

Avoiding DSBs

The CRISPR-SpCas9 system has originally the role of disrupting invading nucleic acids, for which DNA cleavage is essential. However, when employed as a genome editor, relying solely on knock-out (KO) is insufficient to address the diverse scenarios of mutation correction. Moreover, the outcome sequence after the action of NHEJ is not precisely predictable, making it impossible to guarantee the restoration of a wt sequence. DSBs themself come with adverse consequences, including large deletions, rearrangements, chromosomal truncations, translocations, chromothripsis and chromosome loss[82–86]. With these important implications, the development of tools that avoid DSBs has become essential. To bring some remarkable examples, Cas9 is exploited as a DNA-binding domain, when both HNH and RuvC are inactivated. The dead Cas9 (dCas9) in CRISPR activation[87]

and CRISPR interference[88] (CRISPRa/i) technologies are fused to transcriptional domains which increase or repress gene transcription, respectively. Switching to epigenome manipulation, dCas9 can also be combined with effectors able to induce histone methylation[89] or acetylation[90], allowing inheritable gene silencing or activation. CRISPR tools aimed to modify the DNA without DSBs were successfully developed, most of all base and prime editors. In 2019 the first prime editor (PE) was presented as a tool able to precisely install, delete, or replace specific DNA sequences without causing unintended mutations[91]. The PEs are a nicked version of Cas9 (nCas9) fused to a reverse transcriptase, which is guided to the target region by a prime editor gRNA (pegRNA). The pegRNA contains at the 3' the RNA template for the intended mutation, which is synthesized as DNA by the reverse transcriptase and integrated into the genome[91]. Originally, PEs showed low efficiency in mammalian cells but, given its potential, the technology was incredibly improved[92,93], demonstrating to the scientific community how precise DNA mutation could be helpful and safe for genome editing-based therapy[94].

A critical step in exploring DSBs-free genome editing was the introduction of base editors (BE), published by the Liu laboratory. BEs are a revolutionary tool for making precise point mutations in a target DNA sequence without the need for DSBs, donor templates, or HDR[95–97]. They consist of a catalytically impaired CRISPR-Cas nuclease, which is fused to a DNA deaminase, making them able to perform single nucleotide mutation in a targeted manner. More in detail, after the Cas protein binding, the sgRNA anneals the DNA, forming a single strand DNA R-loop[98] in which the PAM-distal nucleotides are affected by the deamination[96,99]. The nucleotide positions within the R-loop that support efficient base editing outcomes define the activity window. For first-generation BEs using SpCas9, this activity window spans protospacer positions 4–8. Nevertheless, it can be influenced by the DNA state, including chromatin architecture, which may vary depending on the genomic locus or cell type[99].

There are two types of base editors: cytosine base editors (CBE) and adenine base editors (ABE), where CBEs can convert C>T, while ABEs mutate A>G (**Fig. 8**). The deamination of C and A generates uridine and inosine respectively, and to achieve stable base editing outcomes, the unedited strand must be replaced to incorporate the corresponding A and C complementary nucleotides opposite the uridine or inosine[96].

CBEs were the first to be presented, and use cytidine deaminases fused to the impaired Cas9 to convert cytosines to uracils, read by polymerases as thymines[96] (**Fig. 8a**). In this work, Komor and colleagues presented three CBEs generations, built one on the other, from BE1 to BE3[96]. Thanks to additional linker optimization and fusion of a second UGI the BE3 activity was improved and the BE4 version was generated[100], while the implementation of a NLS and codon optimization led to BE4max[99,101].

The other main family is ABEs, introduced in 2017. Differently from cytidine deaminases, there are no adenosine deaminases able to use DNA as substrate in nature[95,97], so the tRNA adenosine deaminase (TadA) was evolved to catalyze the deamination on DNA, generating the first version of mutated TadA (TadA*). The fusion of mutated TadAs and nCas9 gave birth to ABEs able to perform targeted A>G conversion since the adenine deamination forms inosine, which is recognized as guanine during DNA replication. The first ABE version was ABE7.10, the result of seven rounds of evolution and engineering of the first ABE attempt, TadA-dCas9[97,102]. ABE7.10 was obtained by combining nCas9 and two TadAs, a wt, and a mutated version, to create heterodimers[103]. ABE7.10 was demonstrated to induce A>G point mutations in living cells with high efficiency and a specific editing window spanning from position 4 to 9 within the protospacer[97]. Compared to CBEs, no specific cellular mechanisms can efficiently excise inosine, so that the adenine base editing product results to be pure, and no structure similar to UGIs is needed in ABEs. Multiple versions were developed to diversify the use of ABEs, by improving the on/off target ratio, expanding the range of selectable PAM sequences, and compacting the ABE construct[104]. ABEmax achieved improved editing efficiency by increasing the number of NLS without altering the editing window[101]. On the other hand, xABE[105], VQR-ABE[106], and NG-ABE[107] saw improvements in both editing efficiency and the editing window by broadening the PAM requirements. Furthermore, the directed evolution of TadA (TadA*) also contributed to enhancing the editing efficiency, as demonstrated in the case of ABE8e[108], whose size results to be sensibly reduced thanks to the elimination of the wt TadA, since it was observed it is not required for ABE activity[109].

Base editing tools present limitations, primarily characterized by restricted possible base pair conversions and targeting sites, which have induced significant modifications for therapeutic applications. Lately, the toolbox of base editors has expanded to enable more conversions, among which C>R transition[110] and A>Y transversion[111,112], thereby extending the capability to address a broader range of point mutations, constituting the predominant category among human pathogenic genetic variants[99]. Early constraints on targeting range were addressed by utilizing Cas9 variants with relaxed PAM requirements[113]. Base editors show unintended edits, including bystander and off-target events. Bystander editing is described as the modification of nucleotides other than the target ones[99], but they can also present undesired conversions in other genomic regions. These have been tackled through deaminase engineering[114], PAM-relaxed Cas9 variants, and optimized gRNA design[102]. Notably, delivering base editors as mRNA or protein:gRNA complexes mitigates off-target effects[102,115]. These modifications collectively aim to address key limitations and advance the therapeutic potential of base editing technologies. Furthermore, base editing holds the advantage of introducing modifications in both actively dividing and non-dividing cells,

providing a clear advantage over HDR, which is constrained to cells undergoing division[116,117].

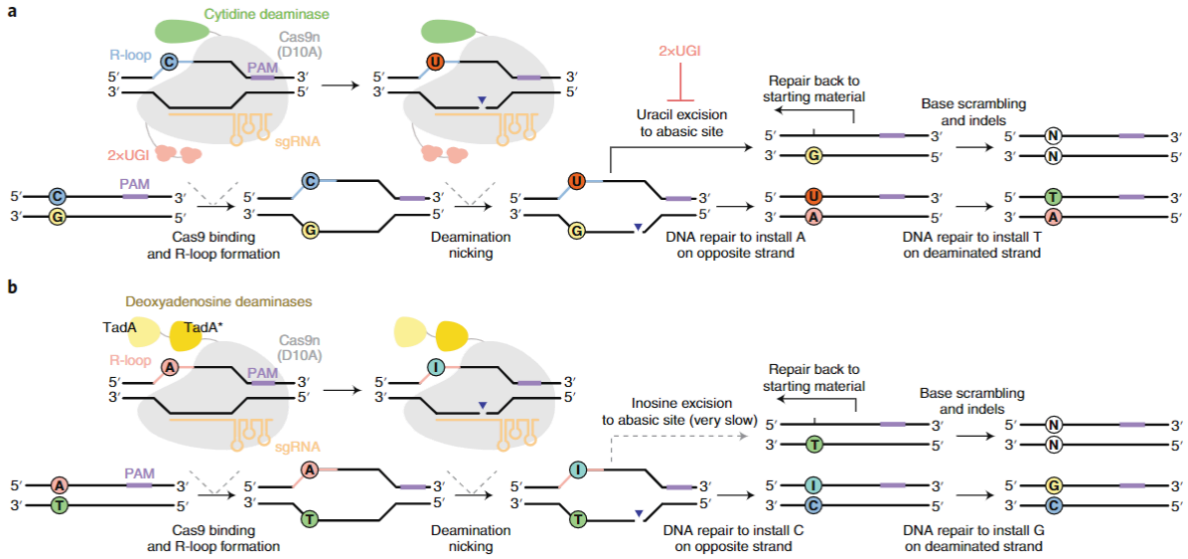

Figure 8. Base editing with CBEs and ABEs. *(**a**) CBEs install C>T substitutions using impaired Cas9s fused to cytidine deaminases and UGIs. (**b**) ABEs install A>G point mutations using a fusion of Cas9 nickase (or dCas9) and TadAs. Adapted from Anzalone et al 2020.*

High-fidelity Cas proteins generation

One of the key challenges associated with an RNA-based system such as CRISPR-Cas9 lies in off-target effects. The Cas9 can tolerate several base pair mismatches between its sgRNA and the target DNA, resulting in incorrect cleavages at sites that do not perfectly match the spacer[80,118]. In the pursuit of improving gene-editing towards more specific technologies, the generation of high-fidelity Cas proteins, especially SpCas9, represented a crucial step. High-fidelity Cas9 versions derive from the modification of the wt sequence through amino acid substitutions obtained after rational engineering and direct evolution. In 2016 the first examples came out, SpCas9-HF1[119] and eSpCas9[120], where contacts with the phosphodiester backbone are weakened in the presence of mismatched targets, while in HypaCas9 the regulation of the HNH domain is stronger to obtain a higher genome-wide precision[121]. xCas9 is a high-fidelity variant evolved to obtain a systematic expansion of the PAM recognition range, broadening its targeting scope while mitigating off-target effects[105]. For the generation of the Cas9_R63A/Q768A76 variant, Bratovic and colleagues identified and mutated arginine residues along the bridge helix of SpCas9 that mediate mismatch

sensitivity[122]. The recently published SuperFi-Cas9 is a variant developed through rational design that introduces mutations to the non-specific DNA-binding region, substantially minimizing undesired interactions[123]. Additionally, the evoCas9[124] and Sniper-Cas9[125] variants integrate directed evolution strategies, endowing them with enhanced fidelity profiles[126].

While these and other high-fidelity Cas9 variants have been successfully developed, they have certain limitations. Specifically, the actual quest is finding high-fidelity Cas9s able to maintain high editing levels, since the generation of specific variants often comes at the cost of a loss in cleavage activity. Moreover, research indicates that using a lower dose of Cas9 can enhance specificity, yet the efficacy of inefficient high-fidelity Cas9 variants may be compromised with transient delivery or in alternative formats (e.g., RNP), particularly in therapeutic applications *in vivo* when the input dose is low.

Delivery of CRISPR-Cas systems

The *in vivo* delivery of genome editing components represents a pivotal challenge of contemporary gene therapy. Viral vectors are strategically used for this purpose, given their natural ability to deliver genetic material into cells. Among these vectors, adenoviral vectors, AAV, and lentiviral vectors have emerged as efficient delivery tools both for bench and bedside applications[127]. Adenoviral vectors, characterized by their episomal persistence, boast a considerable cargo capacity of up to 36kb. However, this is counterbalanced by the drawback of provoking a robust innate immune response, relegating them predominantly to applications such as novel vaccines and cancer therapies, where the immune response can be strategically advantageous[128]. In contrast, lentiviral vectors have found their niche in clinical applications for *ex vivo* gene therapy[129,130]. These vectors can accommodate transgenes of up to 8kb and, in a departure from adenoviral vectors, stably integrate into the host genome. Notably, lentiviral vectors can also be engineered as integrase-deficient variants to bypass integration, expanding their versatility in diverse gene therapy contexts[131]. This unique feature renders lentiviral vectors especially suitable for applications where stable genomic integration is either desired or, conversely, should be avoided. AAVs, at the forefront of *in vivo* gene transfer, exhibit a remarkable safety profile, they do not physiologically integrate into the target genome, and also, the availability of diverse serotypes enables precise targeting of specific tissues or cell populations, augmenting their appeal for therapeutic applications[132]. However, the primary limitation of AAVs lies in their transgene cargo capacity, capped at approximately 4kb. This limitation poses a challenge when attempting to package the complete editing machinery, comprising the Cas9 coding sequence (approximately 4.1kb for SpCas9) and the sgRNA expression construct under the

U6 promoter, necessitating a thoughtful consideration of regulatory sequences for transcription termination and processing.

Alternative strategies for cell delivery have been explored, including the direct delivery of Cas9 protein via electroporation or microinjection of the Cas9/sgRNA RNP complex. This approach offers the advantage of transient Cas9 expression, potentially reducing the off-target activity associated with prolonged exposure to exogenous genome editing reagents[133]. However, direct RNP delivery faces challenges, including the need for purified Cas9 protein and the limitations of current *in vivo* transfection methods, restricting its application primarily to *ex vivo* settings and embryo electroporation or microinjection in animal models.

In the need for new technologies to effectively target cells *in vivo*, lipid nanoparticles (LNP) were developed. These non-viral systems, while less toxic and immunogenic than their viral counterparts[134], have limitations related to low biodistribution and inefficient targeting and penetration of tissues beyond the liver, emphasizing the need for further advancements to broaden the scope of efficient delivery to diverse tissues and cell types. Ultimately, the choice between viral and non-viral delivery systems critically influences the type of Cas9 cargo that can be effectively delivered. While viral vectors are constrained by the nature of their genome, LNPs offer versatility, accommodating diverse cargos in the form of RNA, DNA, or protein[135]. The efficacy of both viral and non-viral vehicles, however, remains constrained by the molecular size of the therapeutic agent, prompting ongoing exploration and innovation toward the development of smaller, yet equally potent, genome editing systems.

Expanding the targetable genome fraction: modify the PAM

The use of wt SpCas9 limits its use to targets containing the 5'-NGG PAM. To expand the applicability and precision of CRISPR-Cas9 genome editing, one of the first SpCas9 optimizations was the modification of the recognized PAM. The first successful attempt was achieved by Kleinstiver *et al*. through random mutagenesis of the PI domain. This approach generated the variants VQR (5'-NGA PAM), EQR (5'-NGAG PAM), and VRER (5'-NGCG PAM), effectively shifting the consensus from 5'-NGG[136]. In 2018, Hu *et al*. used phage-assisted continuous evolution (PACE) to develop xCas9, which demonstrated the ability to recognize 5'-NG, even with some preferences at the third nucleotide position[105]. Concurrently, Nishimashu *et al*. designed SpCas9-NG, a relaxed SpCas9 variant that recognized a 5'-NG consensus PAM with weaker recognition of 5'-NA PAMs. SpCas9-NG exhibited superior recognition of sequences within the 5'-NG motif compared to xCas9, demonstrating greater indel formation and base editing in human cells[107]. Trying to get rid of

all PAM requirements, Walton *et al*. conducted further structure-guided mutagenesis of the VRQR variant to develop a near PAM-less Cas9. This approach, which involved a sequential screening of mutations to key residues impacting PAM recognition, brought to the generation of SpG and SpRY. SpG recognizes a consensus 5'-NG PAM and has been demonstrated to outperform xCas9 for all 5'-NGNN sequences and SpCas9-NG for specific 5'-NGNN sequences. On the other hand, the SpRY variant recognizes a consensus 5'-NR PAM, with a less-preferred recognition of an 5'-RY PAM, presenting the most relaxed PAM preference observed thus far[113]. However, owing to its relaxed PAM preferences, the SpRY variant exhibited a higher tendency for off-targeting compared to SpCas9, albeit based on a limited dataset.

The development of these engineered Cas9 proteins with different PAM requirements is crucial as it enables the scientific community to target a more diverse range of genomic sequences, allowing for more precise genetic manipulations.

Developing and finding new CRISPR-Cas9 systems

The identification of Cas9 proteins with optimized features, consisting in reduced molecular size, precision, and flexible PAM recognition sequences, is pivotal for the progression of genome editing technologies for gene therapy applications. To address this challenge, researchers are actively exploring novel CRISPR-Cas9 systems possessing these specified traits. The first attempt to overcome the SpCas9 size problem was the 1122 aa-long St1Cas9[38,43], isolated from *Streptococcus thermophilus*. However, the PAM requirements are so strict (5′-NNRGAA) it is difficult to employ St1Cas9 for many applications. In the pursuit of smaller yet equally potent Cas9 proteins, the Zhang lab adopted a distinctive approach[137,138]. They analyzed over 600 Cas9 orthologs and identified two distinct groups: longer orthologs, approximately 1350 aa, including SpCas9, and shorter orthologs, around 1000 aa. Among the shorter orthologs, only *Staphylococcus aureus* Cas9 (SaCas9, 1053 aa) exhibited cleavage activity in mammalian cells. SaCas9 demonstrated indel generation with comparable efficiency to SpCas9, prompting the research group to concentrate on characterizing SaCas9 for subsequent *in vivo* studies. The smallest Cas9 to date is *Campylobacter jejuni* Cas9 (CjCas9), with 5'-NNNVRYAC PAM[139], and it was efficiently used *in vivo* via single AAV vector[139–141]. From *Neisseria meningitidis* were isolated NmCas9[142,143] (1081 aa) and Nme2Cas9[144] (1082 aa), which identify a 24 nt protospacer, suggesting a potentially higher level of specificity compared to Cas9s characterized by ~20 nt spacer[138]. In direct comparisons with SpCas9 in human cells, lower off-target mutagenesis was evident; however, a reduction in on-target activity levels was also observed[145]. To conclude this short

overview, SauriCas9[146] from *Staphilococcus auricularis* was characterized since its PAMs, 5'-NNGG, together with its compact size of 1061 aa, being able to target the same sequences of SpCas9 with a single AAV vector. Thanks to their intrinsic attributes, all of these Cas9s can fit in an AAV vector together with their sgRNA and are prone to reduced off-target effects.

Aim of the thesis

In the last ten years, the CRISPR-based technologies were significantly improved but, despite the approval of the first therapeutics using CRISPR, various hurdles persist. The main challenges, especially regarding SpCas9, stem from the lack of balance between on- and off-target efficiency, constraints in targeting range due to PAM requirements, and difficulties in effective delivery caused by the size of Cas nucleases.

This thesis aims to enrich the CRISPR arsenal with new nucleases characterized by beneficial properties such as remarkable activity and precision, relaxed PAM sequence, and small size.

The first aim was the generation of a high-fidelity SpCas9 variant that could preserve its high editing efficacy when delivered as RNP. To guarantee a highly precise mutant, the amino acid substitution was chosen among the evoCas9 screening results, and then, the protein was tested on a panel of genomic loci to evaluate its cleavage efficiency and specificity in human cell lines.

The second and broader aim was finding new CRISPR-Cas9 systems able to satisfy fundamental requirements for a flexible gene editor, searching for new Cas9 proteins in a new unexplored database of bacterial metagenomes. Cas9s were selected based on their size (<1100 aa) and their PAM sequence was identified, to be able to characterize their activity and compare them with the most used Cas9s, such as SpCas9, SaCas9, CjCas9, and Nme2Cas9.

Together with the multitude of Cas tools already available, the Cas9 proteins described in this work build a set of DNA editors, which is constantly developing to target a growing number of conditions.

Part I - Development of rCas9HF, an optimized SpCas9 high-fidelity variant

High-fidelity Cas9 variants, required to control off-target effects, are characterized by a trade-off between activity and specificity. In this work[1], a novel SpCas9 variant, rCas9HF (K526D mutant), was generated, having a favorable specificity profile while keeping activity levels comparable to the wt SpCas9 when delivered as RNP. The study reports a comparative analysis of wt SpCas9, rCas9HF, and HiFi Cas9 (R691A mutant)[147] in various genomic loci, revealing diverse efficiency and precision depending on the target locus. The efficacy of both high-fidelity variants is also demonstrated in editing primary human CD34+ blood stem cells, and in exploiting HDR to repair DSBs. The implementation of rCas9HF enriches the CRISPR toolbox, providing alternative editing solutions.

Results

In this section, references to figures and supplementary are referred to
Pedrazzoli, E. et al. An optimized SpCas9 high-fidelity variant for direct protein delivery. Mol. Ther. 31, 2257–2265 (2023)[1].

Our research for a high-fidelity SpCas9 with a high on/off target ratio was based on our and other groups observation that, even though evoCas9 is the most precise nuclease compared to other SpCas9 derived variants, the editing efficiency is proportionally reduced[148,149]. This feature was also confirmed when evoCas9 was delivered as RNP, compared with WT SpCas9 and HiFi Cas9[147], currently the only efficient high-fidelity Cas9 when exploiting this kind of transfer (**Fig. 1A**).
Thus, the amino acid substitutions resulting from evoCas9 evolution[124] were exploited to find better-performing high-fidelity Cas9s in terms of RNP genome editors. Among these, the K526E mutation was identified as the most effective for a high-fidelity variant maintaining the WT SpCas9 editing efficiency (**Fig. 1B**). However, the single K526E mutant showed editing activity sensibly lower than the WT SpCas9 when targeting endogenous loci (**Fig. 1C**), so all possible amino acid substitutions were tested in K526 position (**Fig. 2A**).
From this analysis, the K526D variant emerged as similarly precise to K526E (**Fig. 2A and Fig. S1**), leading to further characterize this mutant, which was named rCas9HF (recombinant Cas9 high-fidelity). After the Cas9s purification, U2OS cells were electroporated using WT SpCas9, HiFi Cas9, and rCas9HF RNPs to evaluate their editing performance on a panel of 15 genomic loci (**Fig. 2B**). rCas9HF showed similar or better

activity than HiFi Cas9 in 10 loci (*BCR*, *CCR5*, *EMX1*, *FANCF*, *HBB*, *HBG1*, *IL2RG*, *TRAC*, *TRBC*, *ZSCAN2*), and comparable editing to WT SpCas9 in 6 loci (*BCR*, *CCR5*, *EMX1*, *FANCF*, *IL2RG*, *TRAC*). Starting from these positive results, 7 target sites associated with known off-target sites were selected to evaluate the specificity of the variants. The RNP delivery naturally impairs unspecific cleavage, thanks to the rapid RNP clearance in the cell[133,150], so it was necessary to analyze the resulting indels through deep-sequencing, in order to appreciate differences among the variants' activity. As shown in **Figure 2C**, on-target editing patterns are confirmed in all loci, where WT SpCas9 is more active than the two high-fidelity proteins, while, as expected, WT SpCas9 produced more off-targets (**Fig. 2D**). In 3 out of 5 loci, HiFi Cas9 demonstrated higher efficiency than rCas9HF, while at the *CCR5* locus rCas9HF produced 35% of indels and HiFi Cas9 resulted in only 10%. Looking at the off-target cleavage, rCas9HF showed to be more precise than HiFi Cas9 (**Fig. 2D**), especially when evaluating the off/on target ratio derived by the activity of the high-fidelity variants (**Fig. S2A**), where it is possible to appreciate how rCas9HF is significantly more precise at *CCR5*, *HBG,* and *HEKsite4*.

Since the off-target analysis at specific sites represents a partial aspect of Cas9 variants' properties, a genome-wide off-target analysis (GUIDE-seq)[151] was carried out after RNP-based electroporation to target *FANCF*, *VEGFAsite2,* and *VEGFAsite3* with guides reported to cause strong off-targets[124]. rCas9HF and HiFi Cas9 showed to be more precise, having a higher percentage of on-target reads in all the loci compared to WT SpCas9 (**Fig. 2E**). The two variants performed similarly, counting around the 40% of total reads for on-target at *FANCF* locus, 10% for *VEGFAsite2*, and 80% for *VEGFAsite3*, with a lower number of unwanted cleavages (**Fig. S3A**). Interestingly, the three variants demonstrated to have a different pattern of off-target sites (**Fig. S4**).

We evaluated rCas9HF, HiFi Cas9, and WT SpCas9 editing in HSPCs CD34+ at loci with a therapeutic interest (*CCR5*, *HBB* and *HBG1*). While on-target efficiency was similar (**Fig. 3A**), the high-fidelity variants showed reduced off-target cleavages (**Fig. 3B** and **Fig. S2B**). Notably, rCas9HF demonstrated diverse editing and specificity, emphasizing the need for careful SpCas9 variant selection.

Finally, high-fidelity variants, especially rCas9HF, outperformed wt SpCas9 in HDR using single-strand oligonucleotides. Across benchmark loci, rCas9HF consistently produced more precise repairs with lower indels generation (**Fig. 4**), in particular in the *CFTR* locus.

These results highlight the potential of rCas9HF RNP for precise genome editing through NHEJ and HDR both in cell lines and primary cells, offering improved outcomes in clinically relevant genomic sites.

Molecular Therapy
Original Article

An optimized SpCas9 high-fidelity variant for direct protein delivery

Eleonora Pedrazzoli,[1,7] Andrea Bianchi,[1,5,7] Alessandro Umbach,[1] Simone Amistadi,[1,6] Mégane Brusson,[2] Giacomo Frati,[2] Matteo Ciciani,[1] Kalina Aleksandra Badowska,[3] Daniele Arosio,[4] Annarita Miccio,[2] Anna Cereseto,[1,8] and Antonio Casini[3,8]

[1]Department CIBIO, Laboratory of Molecular Virology, University of Trento, Via Sommarive 9, 38123 Trento, Italy; [2]Imagine Institute, Laboratory of Chromatin and Gene Regulation During Development, Université de Paris, INSERM UMR 1163, Paris, France; [3]Alia Therapeutics, 38123 Trento, Italy; [4]Biophysics Institute, National Research Council of Italy, 38123 Trento, Italy

Electroporation of the Cas9 ribonucleoprotein (RNP) complex offers the advantage of preventing off-target cleavages and potential immune responses produced by long-term expression of the nuclease. Nevertheless, the majority of engineered high-fidelity *Streptococcus pyogenes* Cas9 (SpCas9) variants are less active than the wild-type enzyme and are not compatible with RNP delivery. Building on our previous studies on evoCas9, we developed a high-fidelity SpCas9 variant suitable for RNP delivery. The editing efficacy and precision of the recombinant high-fidelity Cas9 (rCas9HF), characterized by the K526D substitution, was compared with the R691A mutant (HiFi Cas9), which is currently the only available high-fidelity Cas9 that can be used as an RNP. The comparative analysis was extended to gene substitution experiments where the two high fidelities were used in combination with a DNA donor template, generating different ratios of non-homologous end joining (NHEJ) versus homology-directed repair (HDR) for precise editing. The analyses revealed a heterogeneous efficacy and precision indicating different targeting capabilities between the two variants throughout the genome. The development of rCas9HF, characterized by an editing profile diverse from the currently used HiFi Cas9 in RNP electroporation, increases the genome editing solutions for the highest precision and efficient applications.

INTRODUCTION

Encouraging results from experimental clinics demonstrate the high potential of the CRISPR technology to treat genetic diseases and cancer.[1,2] Ribonucleoprotein (RNP) delivery is emerging as the method of choice for gene substitution through homology-directed repair[3,4] or targeted sequence disruption as in γ-globin reactivation for the *ex vivo* treatment of β-thalassemia and sickle cell diseases.[1,5–7]

The rapid clearance of RNPs via cellular proteolysis is instrumental in limiting the off-target accumulation produced by Cas9 long-term residency.[8,9] Nonetheless, off-target cleavages are best controlled by using high-fidelity Cas9 variants derived from *Streptococcus pyogenes* Cas9 (SpCas9). High-fidelity Cas9s were generated by either rational protein engineering[10–13] or directed evolution approaches, as for evoCas9,[14] resulting in limited tolerance to mismatches between the guide RNA and DNA substrates.[14–16] A comparative analysis among the high-fidelity variants and the original wild-type SpCas9 (WT SpCas9) demonstrated a general trade-off between Cas9 activity and specificity resulting in an inverse correlation between efficiency and precision.[17,18] The reduced editing efficacy of the high-fidelity SpCas9s can be compensated by sustained protein levels through long-term transgene expression via plasmid or viral vector delivery.[19] Consequently, the fast protein-RNA decay following RNP electroporation is not compatible with the protein levels needed to obtain efficient editing with high-fidelity SpCas9s. In fact, even though a large number of high-fidelity variants have been produced, so far a single mutant, R691A (known as HiFi Cas9), is used in *ex vivo* RNP electroporation.[13] Nonetheless, the editing efficacy and precision varies along the genome,[14] and thus diverse variants responding to diverse editing requirements are needed to reach the highest number of genomic sites through CRISPR editing. To enlarge the spectrum of high-fidelity variants compatible with RNP delivery, we built on the protein engineering work that led to the identification of evoCas9.[14] evoCas9 is characterized by four amino acid substitutions (M495V, Y515N, K526E, R661Q) and shows the highest specificity among all the high-fidelity SpCas9 variants so far developed.[17,18] Nevertheless, its increased specificity is paralleled by a significant loss of activity on several target sites due to a trade-off between activity and specificity.[17,18] Here, we identified the K526D mutant, which we named rCas9HF (recombinant high-fidelity Cas9), showing a favorable

Received 8 December 2022; accepted 7 March 2023;
https://doi.org/10.1016/j.ymthe.2023.03.007.

[5]Present address: Gene Therapy Program, Dana Farber/Boston Children's Cancer and Blood Disorder Center, Harvard Medical School, Boston, MA 02115, USA

[6]Present address: Imagine Institute, Laboratory of Chromatin and Gene Regulation During Development, Université de Paris, INSERM UMR 1163, Paris, France

[7]These authors contributed equally

[8]These authors contributed equally

Correspondence: Anna Cereseto, Department CIBIO, Laboratory of Molecular Virology, University of Trento, Via Sommarive 9, 38123 Trento, Italy.
E-mail: anna.cereseto@unitn.it
Correspondence: Antonio Casini, Alia Therapeutics, 38123 Trento, Italy.
E-mail: antonio.casini@aliatx.com

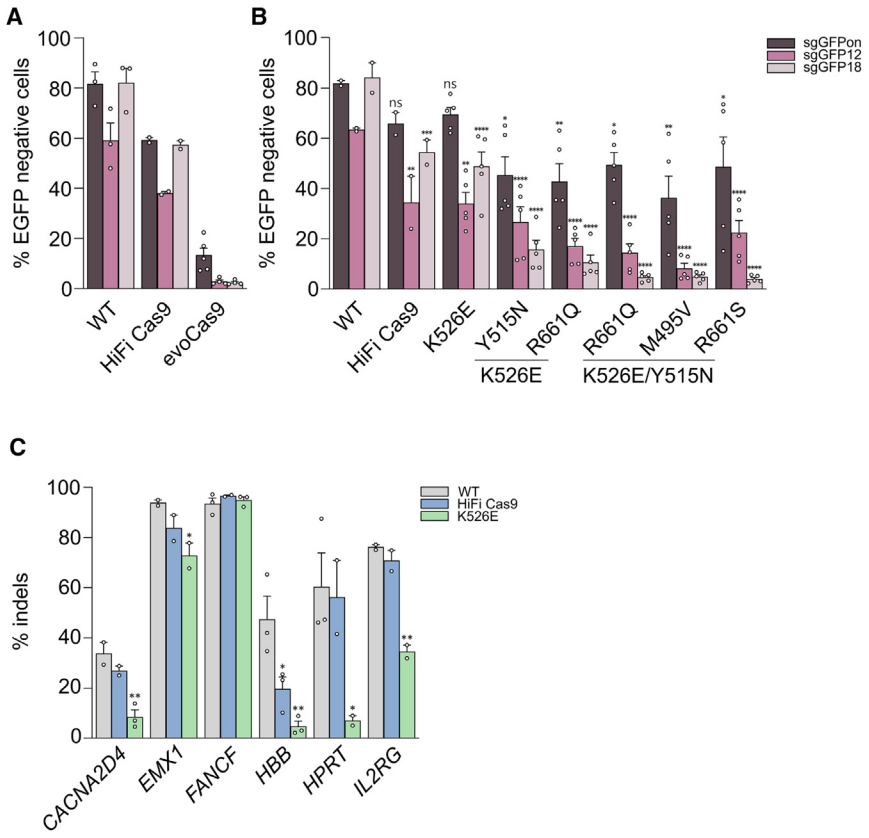

Figure 1. Editing activity and precision of RNP from SpCas9 high-fidelity variants

(A and B) HEK293T cells stably expressing EGFP were lipofected with the indicated RNPs together with sgRNA perfectly matching the target (sgGFPon) or containing a single mismatch at different positions along the spacer as indicated (sgGFP12 and sgGFP18: mismatched nucleotides in positions 12 and 18 counting from the PAM, respectively). Loss of fluorescence was measured by cytofluorimetry 7 days post-lipofection. In (B), statistical significance was assessed using one-way ANOVA, comparing each mutant with WT SpCas9 separately for sgGFPon, sgGFP12, and sgGFP18. (C) Editing activity (percentage of indels) measured by sequencing (tracking indels by deconvolution [TIDE]) at the indicated endogenous genomic loci in U2OS cells 3 days after electroporation of WT SpCas9, HiFi Cas9, or K526E RNPs. Statistical significance was assessed using one-way ANOVA, comparing HiFi Cas9 and K526E with WT SpCas9. Data reported as mean ± SEM for n ≥ 2 biologically independent replicates.

specificity profile, while preserving near-WT levels of activity when delivered as an RNP through electroporation. An in-depth comparative analysis between HiFi Cas9 and rCas9HF in several loci by targeted deep sequencing analysis demonstrated that the two variants have different efficiency and precision depending on the target genomic locus, thus providing alternative editing solutions and enriching the CRISPR toolbox. As a proof of concept, we challenged the two high-fidelity variants by editing a set of loci in primary human CD34+ blood stem cells including the γ-globin promoter, which is currently used as therapeutic target for sickle cell disease and is associated with a known off-target site generated with RNP editing.[7] Finally, we comparatively evaluated their efficacy through RNP delivery in gene substitution experiments using donor DNAs to elucidate the advantages of high-fidelity variants in gene editing involving HDR repair.

RESULTS

Identification of SpCas9 mutants suitable for RNP delivery

Among the SpCas9 high-fidelity variants so far developed, evoCas9 is the most precise, even though this is paralleled by low editing efficiency, especially on specific target sites.[14,17,18] Since evoCas9 has been tested through endogenous expression via plasmid or viral vector delivery, we evaluated its efficacy following delivery as an RNP. We used a lipofection protocol reported before[9] to evaluate the on- and off-target activity of the evoCas9 RNP by targeting EGFP expressed in HEK293T reporter cells.[14] For the on target, we used a single guide RNA (sgRNA) fully matching the EGFP coding sequence, while the off targets were measured with two sgRNAs containing single mismatches at different positions along the same spacer. Despite observing high precision with the off-target sgRNAs (less than 5% off-target cleavages with evoCas9 as opposed to more than 40% with both WT SpCas9 and HiFi Cas9), we detected a very low activity with the on-target sgRNA (on average, 6–8 times less than HiFi Cas9 and WT SpCas9, respectively) (Figure 1A). These results confirmed the inverse correlation between efficiency and precision obtained with the high-fidelity SpCas9 variants,[17,18] which turned to be more obvious by using the RNP delivery.[13] Therefore, starting from the protein engineering work that produced evoCas9, we further refined our findings to finely tune the activity and specificity of the mutants for transient expression, aiming at enlarging the repertoire of SpCas9 variants for RNP delivery. Starting from the pool of most promising mutations identified during the development of evoCas9 (M495V, Y515N, K526E, and R661Q/S),[14] we set to identify alternative combinations of amino acid substitutions that could increase the activity of the RNPs while maintaining the high-fidelity properties. Among these selected mutations, K526E was tested individually given the high specificity previously demonstrated.[14] We thus produced recombinant proteins, including a combination of two or three substitutions, and tested them side by side with WT SpCas9 and HiFi Cas9 using our EGFP reporter cells and surrogate off-target models. As shown in Figure 1B, despite recovering part of the cleavage activity, all triple and most double mutants showed lower on-target editing efficacy compared with both HiFi Cas9 and WT SpCas9. Therefore, we picked the single K526E mutant as the best-performing mutant for further comparative analysis. The experiments were performed using

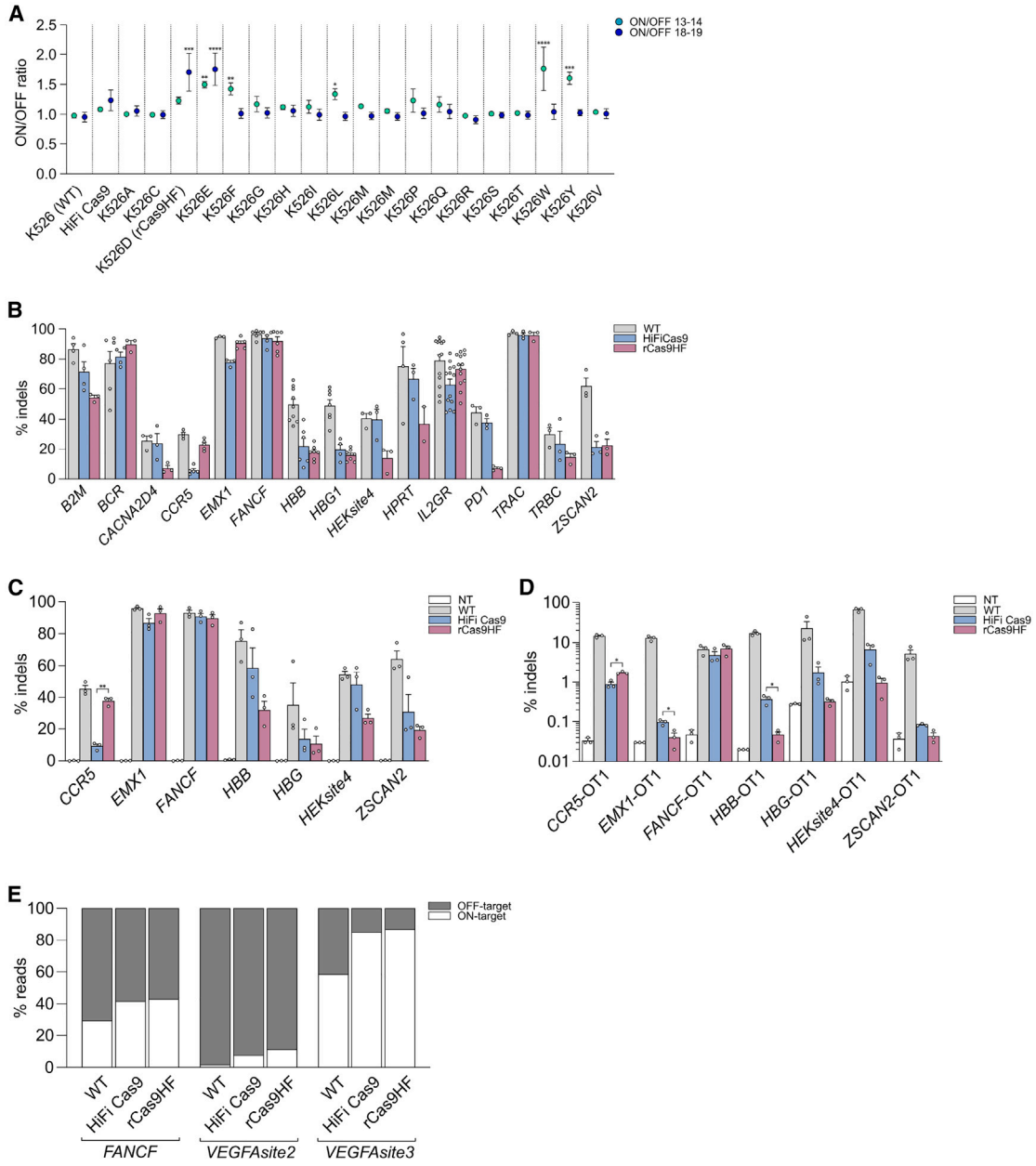

Figure 2. Identification and characterization of an optimized high-fidelity SpCas9 variant

(A) On-/off-target ratios obtained with fully matching sgRNA (on target) over two surrogate off targets (sgGFP13-14 and sgGFP18-19 with double mismatches in position 13-14 or 18-19 of the spacer, respectively) with K526 variants with the indicated amino acid substitutions. The editing was measured following transient plasmid transfection in HEK293-GFP reporter cells. The ratios are obtained using raw data from Figure S1. Statistical significance was assessed using one-way ANOVA, comparing each mutant with WT SpCas9 separately for on/off 13-14 and on/off 18-19. (B) Editing activities (percentage of indels) of WT SpCas9, HiFi Cas9, or rCas9HF RNPs were measured through tracking indels by TIDE analysis (n = 2 for K526D activity on *HPRT* locus). (C and D) Targeted deep sequencing analysis of the on targets (C) and previously validated off targets (D) after electroporation of WT SpCas9, HiFi Cas9, or rCas9HF RNPs in U2OS cells. The off-/on-target ratios calculated from the data in (C) and (D) are reported in Figure S2A. Statistical significance was assessed using paired t test to compare HiFiCas9 and rCas9HF; data reported as mean ± SEM for n ≥ 3 biologically independent replicates. (E) Percentage distribution of GUIDE-seq reads among the on-target and all off-target sites after electroporation of U2OS cells with the WT SpCas9, rCas9HF, and HiFi Cas9 RNPs; GUIDE-seq details are in Figure S3 and Tables S3–S11.

a panel of six endogenous genomic targets and showed that only in the *FANCF* locus did K526E have similar efficiency to the WT and HiFi Cas9, while it was less active in the remaining sites (Figure 1C).

To test whether the loss of editing activity could be recovered while preserving the specificity provided by the modification of the K526 residue, we tested all possible amino acid substitutions in this position. In this set of experiments, we induced higher editing through transient plasmid transfection instead of RNP delivery to enhance the sensitivity for both on- and off-target insertions or deletions (indels), as previously performed.[13] Moreover, in addition to the matched sgRNA for on-target analysis, we used two sgRNAs, each containing two mismatched nucleotides (13-14 and 18-19) in the spacer sequence with increasing distance from the PAM nucleotides, to test the frequency of diverse off-target events. While the on-target activity did not differ significantly among the variants and WT SpCas9, the off-target cleavages varied, showing that the K526D variant was similarly precise as the K526E (Figures 2A and S1). Notably, the editing profile (higher on/off ratios) of both variants was better than the HiFi Cas9 mutant by inducing lower off-target activities (Figures 2A and S1). Given these results, we selected the K526D mutant for further characterization, and we named it rCas9HF.

To investigate the editing performance of the rCas9HF RNP, we measured the indels produced in fifteen endogenous genomic loci. As shown in Figure 2B, rCas9HF had overall similar cleavage efficiency compared with the HiFi Cas9 mutant, thus suggesting that the glutamate-to-aspartate substitution at position K526 re-established near-WT activity. For a thorough comparative analysis among RNP delivery of WT SpCas9, rCas9HF, and HiFi Cas9, we selected the on-target sites from Figure 2B known to be associated with specific off-target sites[7,14,20,21] and analyzed the on- and off-target activity by amplicon sequencing through next-generation sequencing (NGS) (Figures 2C and 2D). The analysis showed that WT SpCas9 is in general more active while producing higher off-target cleavages (Figures 2C and 2D). All the Cas9s had a similar editing efficacy in two sites (*EMX1* and *FANCF*), while for most of the remaining sites (3 out of 5), HiFi Cas9 was more efficient than rCas9HF, except for *CCR5*, where our variant performed considerably better than HiFi Cas9 (Figures 2C and 2D). Nonetheless, the off-target analysis showed that rCas9HF produced less non-specific cleavages (Figures 2C and 2D). For a more direct interpretation of non-specific editing, we graphed the off/on ratios (Figure S2A). Overall, the results showed variable editing between the two high-fidelity variants, even though the off/on target profile is more favorable with the rCas9HF variant, given its higher specificity (Figures 2C, 2D, and S2A).

To further characterize the off-target events, we performed a genome-wide off-target analysis (GUIDE-seq method)[22] following RNP delivery in three genomic loci (*FANCF*, *VEGFAsite2*, and *VEGFAsite3*) that are known to be associated with strong unspecific cleavages.[14] Both the rCas9HF and the HiFi Cas9 variants generated significantly less off-target sites than the WT nuclease as absolute numbers (Figures S3A and S3B) as well as in percentages of on and off reads (Figure 2E). Furthermore, the comparison of the off-target sites captured by GUIDE-seq revealed that in addition to common sites, the engineered and WT variants showed either unique or differently shared off-target sites (Figure S4). Therefore, these results further indicate a more precise editing activity of both high-fidelity variants (slightly better for rCas9HF; Figures 2E, S3A, and S3B) and an off-target profile that is only partially overlapping among the diverse nucleases (Figure S4).

rCas9HF editing in CD34+ primary hematopoietic stem cells

Gene editing in CD34+ hematopoietic stem cells is of particular interest for the development of *ex vivo* gene therapies, with very encouraging recent results from the experimental clinic.[1] We thus evaluated the activity and specificity of rCas9HF, HiFi Cas9, and WT SpCas9 on selected loci with therapeutic relevance in CD34+ stem cells (*CCR5*, *HBB*, *HBG*) through targeted deep sequencing analysis to increase sensitivity. After RNP electroporation in CD34+ cells, the editing produced by WT SpCas9 and the high-fidelity variants was measured at the on-target as well at the associated off-target (OT) sites.[7,14,23] The three nucleases showed similar on-target efficiency (Figure 3A), while the OT cleavages were overall less frequent with the high-fidelity variants (Figure 3B). Additionally, the off/on values calculated from these editing data clearly showed the diverse editing and specificity of the rCas9HF and HiFi Cas9 variants, which might be relevant at targeted therapeutic sites, soliciting a careful evaluation of the SpCas9 variant to use for each specific locus (Figure S2B).

Improvement of editing through homology-directed repair by high-fidelity variants

High-fidelity variants are more efficient than WT SpCas9 in sequence substitution protocols through homology-directed repair (HDR) using donor templates. In fact, compared with WT SpCas9, high-fidelity variants offer the advantage of minimizing the re-editing of the repaired locus as a consequence of their lower tolerance to mismatches between the sgRNA and the integrated donor template carrying silenced mutations.[7,14,23–25] We thus compared the performance of our best high-fidelity mutant, rCas9HF, with HiFi Cas9 and WT SpCas9 toward three benchmark genomic loci: *EMX1*, *HBB*, and *BCR*. In all tested loci, single-strand oligonucleotides (ssODNs) were used as donor DNAs containing a mismatched nucleotide within the range of the sgRNA target sequence (Table S1). Overall, the editing efficacy was similar among WT SpCas9 and the high fidelity variants, except for a lower editing by rCas9HF in one (the *HBB* site) out of three sites (Figure 4A). Nonetheless, the comparative analysis of non-homologous end joining (NHEJ)-generated indels compared with precise HDR showed that, in all cases, our variant produced more precise repairs (Figures 4A and 4B). This result is consistent with former reports demonstrating that increased precision correlate with higher HDR products.[20]

Precise repair through HDR was measured in a therapeutically relevant locus, *CFTR*, mutations of which cause cystic fibrosis and which has been widely tested for the development of therapeutic genome

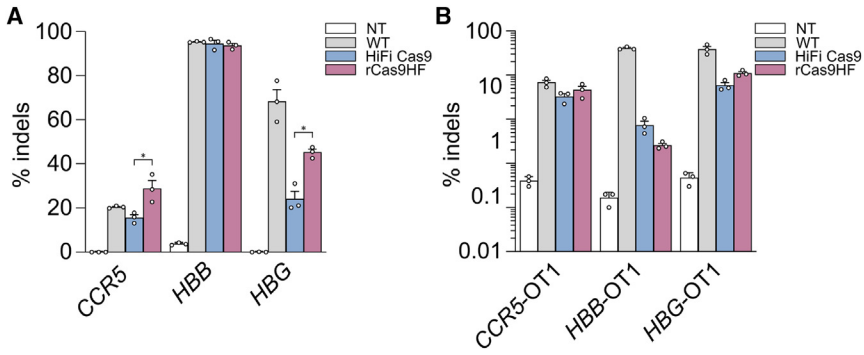

Figure 3. Editing of clinically relevant genomic loci by RNP delivery of high-fidelity SpCas9 variants in CD34+ cells

Primary CD34+ cells were electroporated with WT SpCas9, HiFi Cas9, or K562D RNPs targeting the *CCR5*, *HBG*, and *HBB* loci. Targeted deep sequencing analyses were performed 6 days post-transfection at the on-target sites (A) or at reported identified off targets (B). The off-/on-target ratios for WT SpCas9 and the two SpCas9 high-fidelity variants (HiFi Cas9 and rCas9HF) calculated from (A) and (B) are reported in Figure S2B. Statistical significance was assessed using paired t test; data reported as mean ± SEM for n = 3 biologically independent replicates.

editing approaches.[26] In particular, we focused on two mutations: (1) the most common cystic fibrosis alteration consisting in a small deletion (*CFTR*-ΔF508) and (2) the *CFTR* 1717-1 G>A point mutation in intron 11 resulting in splicing defects.[27] Cells treated with rCas9HF showed less indels than those produced by both WT SpCas9 and the HiFi Cas9 (Figure 4C), which correlated with higher HDR compared with WT SpCas9 or HiFi Cas9 (Figures 4C and 4D). As above (Figures 4A and 4B), increased editing precision through HDR is accompanied by lower NHEJ editing products generated by both high-fidelity variants (Figures 4A–4D).

DISCUSSION

Two fundamental parameters in genome editing are efficiency and precision. CRISPR-Cas9 precision was highly improved through the generation of high-fidelity variants[17] and by reducing intracellular levels of Cas9.[20] One of the most successful methods for transient CRISPR-Cas9 expression is RNP electroporation, which results in the quickest degradation kinetics of the enzymatic complex in target cells (24–48 h persistence)[8,9] compared with mRNA (48–72 h persistence) or plasmid delivery (>72 h persistence). Nevertheless, RNP delivery is hardly compatible with high-fidelity variants,[13] resulting in low editing efficiency as also confirmed by our results (Figure 1). Very likely, the slow cleavage kinetics of the high-fidelity nucleases require longer target residency and higher enzymatic concentrations compared with the original WT SpCas9.[8] In fact, so far, exclusively one high-fidelity mutant, the HiFi Cas9, has been identified to efficiently edit the genome through RNP electroporation and is thus unique for *ex vivo* editing.[13,23]

Nonetheless, the activity and precision of CRISPR-Cas nucleases throughout the genome is not homogeneous[14] (as also demonstrated in this study; Figures 1C and 2B–2E), most likely due to different levels of chromatic compaction across different genomic loci.[28–30] For this reason, an expansion of the editing toolbox is clearly needed to efficiently address all editing requirements. This can be obtained either by engineering CRISPR enzymes with improved properties, such as rCas9HF, or through the discovery of completely undescribed nucleases.[31] Beyond editing through indels generated by NHEJ, another important use of high-fidelity variants through RNP delivery is gene substitution through HDR using DNA donor templates.[4] The development of a high-fidelity SpCas9 nuclease working through RNP delivery, the rCas9HF variant, increases the opportunity for efficient and precise editing on a wider range of targeted loci and for more efficient HDR protocols.

The identification of the rCas9HF in this study was obtained by introducing a single point mutation in the WT nuclease. The amino acid interested by the substitution was originally identified during the development of evoCas9, which contained four mutations (M495V, Y515N, K526E, and R661Q/S) localized in the REC3 domain at protein sites where SpCas9 contacts the RNA:DNA duplex.[14,32] The experimental work, which led to the identification of evoCas9, revealed that, similarly to other high-fidelity variants, the activity profile of the nuclease results from a trade-off between efficiency and specificity, with the highest precision obtained at the expense of nuclease activity.[17,18]

The best compromise to balance precision and efficiency for RNP delivery was obtained by installing a single mutation at position K526, which we found best works with an aspartate in place of the original glutamate present in evoCas9. Based on these results, we hypothesize that the increased activity observed with K526D, compared with K526E, stems from the aspartate's shorter side chain: while both substitutions lead to a charge inversion, K526D is better accommodated within the context of neighboring residues, thus preserving the catalytic activity (see *in silico* analysis in Figure S5). Further structural studies are needed to strengthen the role of the K526E residue as a potential groundwork to further develop high-fidelity variants including other Cas9 orthologs for RNP delivery.

Finally, the immunogenicity of SpCas9 raises concern for its wider use in the clinic, as both humoral and cellular immune responses have been reported in the population.[33] By cross checking the reported SpCas9 immunodominant epitopes[34] with the amino acid region where the K526D substitution is located, we found no overlap, indicating that the mutation should not be able to modulate the immunological properties of the nuclease.

Here, we generated rCas9HF, a high-fidelity variant suitable for RNP delivery that is an alternative to the so-far unique HiFi Cas9. Despite

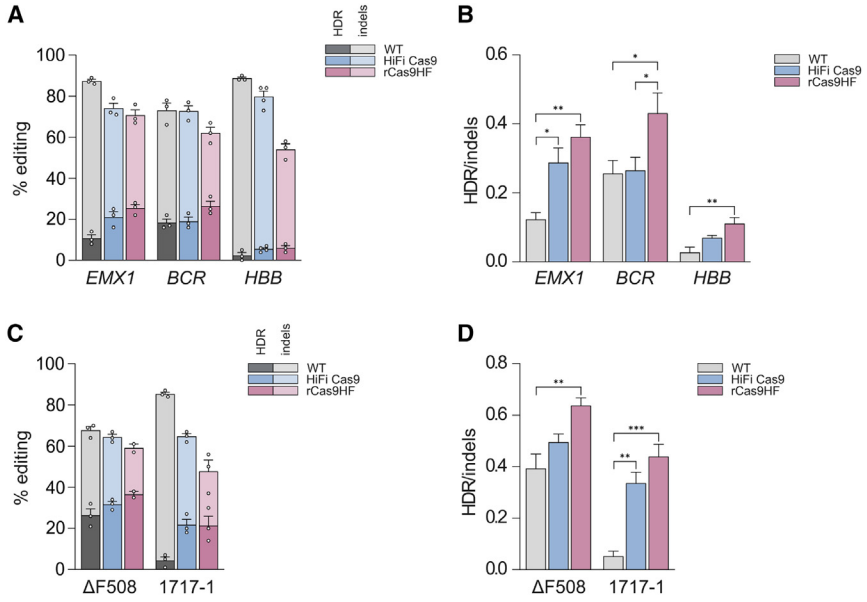

Figure 4. Precise gene correction by HDR using high-fidelity RNPs

(A) U2OS cells were electroporated with WT SpCas9, HiFi Cas9, or rCas9HF RNPs targeting the *EMX1*, *BCR*, and *HBB* loci together with specific ssODN templates (Table S1). Indel formation (NHEJ) and HDR events were quantified at 3 days post-transfection. (B) HDR/indel ratios calculated for the *EMX1*, *BCR*, and *HBB* loci are from (A). (C) Quantification of HDR-mediated correction in the *CFTR* locus targeting two clinically relevant mutations causing cystic fibrosis. HEK293 cells stably expressing mutated *CFTR* minigenes were electroporated with WT SpCas9, HiFi Cas9, or rCas9HF RNP complexes and an ssODN donor. Indel frequencies and HDR levels were analyzed 3 days post-electroporation. (D) HDR/indel ratios calculated for the two *CFTR* loci are from (C). Data reported as mean ± SEM for $n \geq 2$ biologically independent replicates; statistical significance was assessed using one-way ANOVA. Bars in (A) and (C) are superimposed.

the similar editing profile of the two high-fidelity nucleases, the in-depth analysis in primary CD34+ cells demonstrates that in particular for *ex vivo* clinical applications, the two variants can represent alternative tools to optimize editing approaches in terms of activity and precision. This represents a step forward in enriching the genome editing toolbox also by providing further insights on the generation of Cas9 mutants with enhanced targeting precision.

MATERIALS AND METHODS

Plasmids

A derivative of pX330 (Addgene #42230) where the sgRNA cassette has been removed by NdeI digestion was used to express SpCas9 in mammalian cells (pX-Cas9). pX-Cas9 plasmids encoding the different mutants were obtained by site-directed mutagenesis of the WT pX-Cas9 plasmid using the ODNs reported in Table S2. sgRNAs were expressed from a pUC19 plasmid containing a U6-driven sgRNA expression cassette. Desired spacer sequences were cloned as annealed ODNs (Table S1) into a double BbsI site immediately upstream of the sgRNA scaffold according to previously published cloning strategies.[35]

To express SpCas9 in bacterial cells for recombinant protein production, a modified version of the pET-28b-Cas9-His (Addgene #47327) was used. Briefly, an additional SV40 nuclear localization sequence (NLS) was added both at the N terminus and C terminus of the protein using standard cloning techniques, bringing the total number of NLSs to three (1 at the N terminus and 2 at the C terminus), generating pET-28b-NLS-Cas9-2xNLS-His. Corresponding plasmids to purify the different SpCas9 mutants were generated by site-directed mutagenesis using the ODNs reported in Table S2.

WT and 1717-1G>A-mutated *CFTR* minigenes were cloned into pcDNA3 (Invitrogen). The WT pMG1717-1WT minigene was obtained by PCR amplification and cloning of target regions from *CFTR* gene from HEK293T cells using the ODNs listed in Table S2. The pMG1717-1WT plasmid contains full exons 10, 11, 12, and 13, portions of intron 11, and full intron 12. The mutated pMG1717-1G>A minigene was obtained by site-directed mutagenesis of the WT minigene construct using the ODNs listed in Table S2.

Cells

U2OS cells were obtained from ATCC (HTB-96). HEK293 cells stably expressing multiple copies of EGFP (293multiEGFP cells) have been previously described.[14] HEK293/CFTR-ΔF508 were obtained by transduction of a lentiviral vector expressing CFTR-ΔF508[36] (a kind gift of Marianne Carlon, KU Leuven). HEK293/1717-1G>A stably expressing pMG1717-1G>A were produced by stable transfection of the BglII-linearized minigene plasmid in HEK293 cells. Cells were cultured in DMEM (Gibco) supplemented with fetal bovine serum (FBS; 10%, Gibco), glutamine (Gibco), and penicillin/streptomycin (Gibco) and maintained at 37°C in 5% CO_2 humidified atmosphere. 293multiEGFP and HEK293CFTR-ΔF508 culture medium was additionally supplemented with 1 μg/mL puromycin. 500 μg/mL G418 (Thermo Fisher Scientific) was added to HEK293/1717-1G>A culture medium. Selection was removed during transfection experiments.

We obtained human peripheral blood $CD34^+$ cells from patients with sickle cell disease. Samples eligible for research purposes were obtained from the "Hôpital Necker-Enfants malades" Hospital (Paris, France). Written informed consent was obtained from all adult subjects. All experiments were performed in accordance with the Declaration of Helsinki. The study was approved by the regional investigational review board (reference: DC 2014-2272, CPP Ile-de-France II "Hôpital Necker-Enfants malades"). $CD34^+$ cells were purified by immunomagnetic selection with AutoMACS (Miltenyi Biotec)

after immunostaining with the CD34 MicroBead Kit (Miltenyi Biotec). 48 h before transfection, $CD34^+$ cells (5×10^5 cells/mL) were thawed and cultured at 37°C in 5% CO_2 humidified atmosphere in a medium containing StemSpan (STEMCELL Technologies) supplemented with penicillin/streptomycin (Gibco), 250 nM StemRegenin1 (STEMCELL Technologies), and the following recombinant human cytokines (PeproTech): human stem cell factor (SCF) (300 ng/mL), Flt-3L (300 ng/mL), thrombopoietin (TPO) (100 ng/mL), and interleukin-3 (IL-3) (60 ng/mL).

Transfection of 293multiEGFP cells

RNP lipofection was performed 24 hours after seeding 1.5×10^5 293multiEGFP cells in a 24-well plate. Guide RNAs were obtained from IDT as separate chemically stabilized CRISPR RNAs (crRNAs) and trans-activating crRNA (tracrRNA), while WT SpCas9 and its mutants were purified as described (protein purification). RNPs were obtained assembling 6 pmol of previously annealed crRNA:-tracrRNA and 3 pmol of SpCas9 protein of interest, then cells were lipofected using TransIT-X2 (Mirus Bio), according to the manufacturer's protocol.

Plasmid transfections were performed by seeding 10^5 293multiEGFP cells in a 24-well plate the day before transfection. 400 ng of each pX-Cas9 plasmid (WT and mutants) together with 200 ng pUC19-sgRNA plasmids were transfected using the TransIT-LT1 reagent (Mirus Bio) according to the manufacturer's protocol. Cells were collected 7 days post-transfection and analyzed by flow cytometry using FACSCanto (BD Biosciences) to evaluate EGFP knockout.

Protein purification

Recombinant nucleases were purified according to a previously published protocol with minor modifications.[37] Briefly, each pET-28b-NLS-Cas9-2xNLS-His was transformed in Rosetta (DE3) pLysS *E. coli* competent cells (Novagen), and single colonies were inoculated and grown overnight, shaking at 37°C to obtain starter cultures, which were then diluted 1:100 in 1–2 L LB medium. Mass cultures were grown at 37°C until OD600 reached 0.6, they were then transferred at 18°C, and after 30 min, IPTG (final concentration 400 mM) was added to the medium to induce SpCas9 expression. Cultures were left shaking at 18°C overnight. Bacterial pellets were then re-suspended in lysis buffer (20 mM Tris [pH 8], 500 mM NaCl, and 5 mM imidazole) supplemented with lysozyme (0.5 mg/mL), incubated for 20 min at 4°C while shaking and lysed by sonication. After clarification by centrifugation (30,000 × *g*, 20 min, 4°C), the supernatants were mixed with NiNTA agarose beads (Protino, Macherey-Nagel) for initial protein isolation. According to protocol, two further steps of purification were performed: ion-exchange (IEX) chromatography (HiTrap SP FF column, GE Healthcare) followed by size-exclusion chromatography (HiLoad 16/600 Superdex 200 PG, GE Healthcare) using an AKTA Pure 25 FPLC system (GE Healthcare). After the final elution, fractions containing SpCas9 were pooled, concentrated using a centrifugal concentrator (10,000 MWCO) to reach at least 10 mg/mL, and stored in aliquots at −80°C until use.

Electroporation of U2OS cells

RNP electroporation experiments in U2OS and HEK293 minigene-expressing cells were performed using the Lonza 4D-Nucleofector (Lonza). RNP complexes were assembled in a final volume of 5 μL using 120 pmol previously annealed crRNA-tracrRNA (IDT) duplex and 100 pmol purified SpCas9. Complexes were mixed with 2×10^5 U2OS cells re-suspended in 20 μL SE buffer and electroporated using the CM-104 program. Genomic DNA was extracted 48 h post-electroporation to assess editing efficiency.

For RNP and ssODN delivery in HDR experiments, 120 pmol Cas9 protein and 150 pmol assembled crRNA-tracrRNA (IDT) were incubated at room temperature for 10–20 min, mixed with 120 pmol ssODN (IDT) and 120 pmol Alt-R Ca9 Electroporation Enhancer (IDT), and electroporated in a total of 2×10^5 U2OS or HEK293 stably expressing mutated CFTR minigenes in SE buffer (programs CM-104 and CM-130, respectively). ssODN donor sequences are reported in Table S1.

Electroporation of CD34+ cells

sgRNA containing both crRNA and tracrRNA sequences was obtained from Synthego (Table S1). RNP complexes were assembled in a final volume of 2.3–2.8 μL using 180 pmol sgRNA and 90 pmol purified SpCas9. CD34+ cells (2×10^5 cells/condition) were transfected in the presence of 180 pmol Alt-R Cas9 Electroporation Enhancer (IDT). We used the Lonza 4D-Nucleofector (Lonza), the P3 Primary Cell 4D-Nucleofector X Kit S (Lonza), and the CA137 program. After transfection, cells were kept in the same medium for 6 days. Genomic DNA was extracted 6 days post-electroporation, and the editing was assessed by deep-sequencing (Table S1).

Evaluation of genome editing by Sanger sequencing

Genomic DNA was extracted from cell pellets using the QuickExtract DNA extraction solution (Lucigen) according to the manufacturer's instructions. Edited regions were amplified using ODNs listed in Table S1. Indels and HDR events were evaluated by deconvolution of chromatographic traces using the tracking indels by deconvolution (TIDE) software (http://shinyapps.datacurators.nl/tide/)[38] or the Synthego ICE analysis tool (v.2) after Sanger sequencing the purified PCR products (EasyRun Service, Microsynth). Untreated cells were used as negative controls for calculating background modification frequencies.

Targeted deep-sequencing

Selected OT sites for *CCR5*, *EMX1*, *FANCF*, *HBB*, *HBG*, *HEKsite4*, and *ZSCAN2* genomic loci, together with their relative on target, were amplified using the HOT FIREPol DNA Polymerase (Solis BioDyne) from genomic DNA of U2OS and CD34+ cells extracted 48 h after electroporation using WT SpCas9, HiFi Cas9, or rCas9HF. Amplicons were indexed by PCR using Nextera indexes (Illumina) and quantified with the Qubit dsDNA High Sensitivity Assay kit (Invitrogen). OT and on-target amplicons were pooled at a 10:1 M ratio and sequenced on an Illumina Miseq system using an Illumina Miseq Reagent kit v.3 for 300 cycles (2 × 150 bp paired end). The complete primer list used to

generate the amplicons is reported in Table S1. Deep sequencing reads were analyzed using CRISPResso2 (v.2.0.45).[39]

OT evaluation

GUIDE-seq experiments were performed as previously described.[22] Briefly, 2×10^5 U2OS cells were electroporated with the Lonza 4D-Nucleofector (Lonza) DN-100 program using the Cas9-gRNA RNP and adding 50 pmol dsODNs (crRNAs in Table S1); cells treated using WT SpCas9 and double-strand ODNs (dsODNs) were used as negative control. 3 days after transfection, cells were collected, and genomic DNA was extracted using NucleoSpin Tissue Kit (Macherey-Nagel) following the manufacturer's instructions. Using the Focused Ultrasonicator (Covaris), genomic DNA was sheared to an average length of 500 bp. End-repair reaction was performed using NEBNext Ultra End Repair/dA Tailing Module and adaptor ligation using NEBNext Ultra Ligation Module as described by Nobles et al.[40] Amplification steps were then performed following the GUIDE-seq original protocol.[22] Following quantification of the libraries by Qubit dsDNA High Sensitivity Assay kit (Invitrogen), the MiSeq sequencing system (Illumina) was used with an Illumina Miseq Reagent kit (150 cycles, paired-end reads). Sequencing data were analyzed using the GUIDE-seq package (v.1.0.2).[22] Visualization of aligned OT sites is available as a color-coded sequence grid (Figure S3B). GUIDE-seq data are listed in Tables S3–S11.

Statistical analysis

All statistics were calculated using GraphPad Prism. Ordinary one-way analysis of variance (ANOVA) was used in Figures 1, 2, and 4 as described in the captions. HiFi Cas9 and rCas9HF were compared using paired t test. Statistical significance was defined as $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, $^{****}p < 0.0001$, ns, non-significant.

DATA AND MATERIALS AVAILABILITY

GUIDE-seq and targeted deep sequencing data have been deposited at BioProject (https://www.ncbi.nlm.nih.gov/bioproject/) under the accession number BioProject: PRJNA939956. All other relevant data are available from the authors upon request.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.ymthe.2023.03.007.

ACKNOWLEDGMENTS

We thank lab members of Cereseto lab for helpful discussions. This work was supported by the European Union's Horizon 2020 innovation program through the UPGRADE (Unlocking Precision Gene Therapy) project (grant agreement no. 825825), by the Horizon Europe EIC Pathfinder program under grant agreement no. 101071041 (AAVolution), and by the Italian Foundation for Cystic Fibrosis (FFC#3-2019). The authors are grateful to Gino Del Bon for supporting E.P. and A.U.'s fellowship through "progetto Sofia."

AUTHOR CONTRIBUTIONS

A. Casini and A. Cereseto designed the research. E.P. and A.B. performed the protein engineering and genome editing analyses. D.A. analyzed the protein structure. G.F., M.B., and A.M. analyzed the RNP editing in primary cells from sickle cell disease patients. K.A.B. performed genome editing experiments. A.U. and S.A. performed HDR experiments. M.C. performed all the bioinformatic analyses. E.P., A.U., and S.A. performed GUIDE-seq experiments.

DECLARATION OF INTERESTS

A. Casini and A. Cereseto are founders of and hold shares in Alia Therapeutics, a genome editing company. A. Casini is an employee of Alia Therapeutics, and A. Cereseto is a consultant for Alia Therapeutics. A patent application covering the technology disclosed in this manuscript has been filed, and A. Casini and A. Cereseto are listed as inventors.

REFERENCES

1. Frangoul, H., Altshuler, D., Cappellini, M.D., Chen, Y.-S., Domm, J., Eustace, B.K., Foell, J., de la Fuente, J., Grupp, S., Handgretinger, R., et al. (2021). CRISPR-Cas9 gene editing for sickle cell disease and β-thalassemia. N. Engl. J. Med. *384*, 252–260. https://doi.org/10.1056/NEJMoa2031054.

2. Hirakawa, M.P., Krishnakumar, R., Timlin, J.A., Carney, J.P., and Butler, K.S. (2020). Gene editing and CRISPR in the clinic: current and future perspectives. Biosci. Rep. *40*. https://doi.org/10.1042/BSR20200127.

3. Lattanzi, A., Camarena, J., Lahiri, P., Segal, H., Srifa, W., Vakulskas, C.A., Frock, R.L., Kenrick, J., Lee, C., Talbott, N., et al. (2021). Development of β-globin gene correction in human hematopoietic stem cells as a potential durable treatment for sickle cell disease. Sci. Transl. Med. *13*, eabf2444. https://doi.org/10.1126/scitranslmed.abf2444.

4. Ghetti, S., Burigotto, M., Mattivi, A., Magnani, G., Casini, A., Bianchi, A., Cereseto, A., and Fava, L.L. (2021). CRISPR/Cas9 ribonucleoprotein-mediated knockin generation in hTERT-RPE1 cells. STAR Protoc. *2*, 100407. https://doi.org/10.1016/j.xpro.2021.100407.

5. Pavani, G., Fabiano, A., Laurent, M., Amor, F., Cantelli, E., Chalumeau, A., Maule, G., Tachtsidi, A., Concordet, J.-P., Cereseto, A., et al. (2021). Correction of β-thalassemia by CRISPR/Cas9 editing of the α-globin locus in human hematopoietic stem cells. Blood Adv. *5*, 1137–1153. https://doi.org/10.1182/bloodadvances.2020001996.

6. Frati, G., and Miccio, A. (2021). Genome editing for β-hemoglobinopathies: advances and challenges. J. Clin. Med. *10*, 482. https://doi.org/10.3390/jcm10030482.

7. Weber, L., Frati, G., Felix, T., Hardouin, G., Casini, A., Wollenschlaeger, C., Meneghini, V., Masson, C., De Cian, A., Chalumeau, A., et al. (2020). Editing a γ-globin repressor binding site restores fetal hemoglobin synthesis and corrects the sickle cell disease phenotype. Sci. Adv. *6*, eaay9392. https://doi.org/10.1126/sciadv.aay9392.

8. Kim, S., Kim, D., Cho, S.W., Kim, J., and Kim, J.-S. (2014). Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. Genome Res. *24*, 1012–1019. https://doi.org/10.1101/gr.171322.113.

9. Liang, X., Potter, J., Kumar, S., Zou, Y., Quintanilla, R., Sridharan, M., Carte, J., Chen, W., Roark, N., Ranganathan, S., et al. (2015). Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. J. Biotechnol. *208*, 44–53. https://doi.org/10.1016/j.jbiotec.2015.04.024.

10. Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z., and Joung, J.K. (2016). High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. Nature *529*, 490–495. https://doi.org/10.1038/nature16526.

11. Chen, J.S., Dagdas, Y.S., Kleinstiver, B.P., Welch, M.M., Sousa, A.A., Harrington, L.B., Sternberg, S.H., Joung, J.K., Yildiz, A., and Doudna, J.A. (2017). Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. Nature *550*, 407–410. https://doi.org/10.1038/nature24268.

12. Slaymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X., and Zhang, F. (2016). Rationally engineered Cas9 nucleases with improved specificity. Science *351*, 84–88. https://doi.org/10.1126/science.aad5227.

13. Vakulskas, C.A., Dever, D.P., Rettig, G.R., Turk, R., Jacobi, A.M., Collingwood, M.A., Bode, N.M., McNeill, M.S., Yan, S., Camarena, J., et al. (2018). A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells. Nat. Med. *24*, 1216–1224. https://doi.org/10.1038/s41591-018-0137-0.

14. Casini, A., Olivieri, M., Petris, G., Montagna, C., Reginato, G., Maule, G., Lorenzin, F., Prandi, D., Romanel, A., Demichelis, F., et al. (2018). A highly specific SpCas9 variant is identified by in vivo screening in yeast. Nat. Biotechnol. *36*, 265–271. https://doi.org/10.1038/nbt.4066.

15. Liu, R., Liang, L., Freed, E.F., and Gill, R.T. (2021). Directed evolution of CRISPR/Cas systems for precise gene editing. Trends Biotechnol. *39*, 262–273. https://doi.org/10.1016/j.tibtech.2020.07.005.

16. Lee, J.K., Jeong, E., Lee, J., Jung, M., Shin, E., Kim, Y.-H., Lee, K., Jung, I., Kim, D., Kim, S., and Kim, J.S. (2018). Directed evolution of CRISPR-Cas9 to increase its specificity. Nat. Commun. *9*, 3048. https://doi.org/10.1038/s41467-018-05477-x.

17. Schmid-Burgk, J.L., Gao, L., Li, D., Gardner, Z., Strecker, J., Lash, B., and Zhang, F. (2020). Highly parallel profiling of Cas9 variant specificity. Mol. Cell *78*, 794–800.e8. https://doi.org/10.1016/j.molcel.2020.02.023.

18. Kim, N., Kim, H.K., Lee, S., Seo, J.H., Choi, J.W., Park, J., Min, S., Yoon, S., Cho, S.-R., and Kim, H.H. (2020). Prediction of the sequence-specific cleavage activity of Cas9 variants. Nat. Biotechnol. *38*, 1328–1336. https://doi.org/10.1038/s41587-020-0537-9.

19. Lino, C.A., Harper, J.C., Carney, J.P., and Timlin, J.A. (2018). Delivering CRISPR: a review of the challenges and approaches. Drug Deliv. *25*, 1234–1257. https://doi.org/10.1080/10717544.2018.1474964.

20. Petris, G., Casini, A., Montagna, C., Lorenzin, F., Prandi, D., Romanel, A., Zasso, J., Conti, L., Demichelis, F., and Cereseto, A. (2017). Hit and go CAS9 delivered through a lentiviral based self-limiting circuit. Nat. Commun. *8*, 15334–15339. https://doi.org/10.1038/ncomms15334.

21. Martin, R.M., Ikeda, K., Cromer, M.K., Uchida, N., Nishimura, T., Romano, R., Tong, A.J., Lemgart, V.T., Camarena, J., Pavel-Dinu, M., et al. (2019). Highly efficient and marker-free genome editing of human pluripotent stem cells by CRISPR-Cas9 RNP and AAV6 donor-mediated homologous recombination. Cell Stem Cell *24*, 821–828.e5. https://doi.org/10.1016/j.stem.2019.04.001.

22. Tsai, S.Q., Zheng, Z., Nguyen, N.T., Liebers, M., Topkar, V.V., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, A.J., Le, L.P., et al. (2015). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat. Biotechnol. *33*, 187–197. https://doi.org/10.1038/nbt.3117.

23. Dever, D.P., Bak, R.O., Reinisch, A., Camarena, J., Washington, G., Nicolas, C.E., Pavel-Dinu, M., Saxena, N., Wilkens, A.B., Mantri, S., et al. (2016). CRISPR/Cas9 β-globin gene targeting in human haematopoietic stem cells. Nature *539*, 384–389. https://doi.org/10.1038/nature20134.

24. Idoko-Akoh, A., Taylor, L., Sang, H.M., and McGrew, M.J. (2018). High fidelity CRISPR/Cas9 increases precise monoallelic and biallelic editing events in primordial germ cells. Sci. Rep. *8*, 15126. https://doi.org/10.1038/s41598-018-33244-x.

25. Kato-Inui, T., Takahashi, G., Hsu, S., and Miyaoka, Y. (2018). Clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein 9 with improved proof-reading enhances homology-directed repair. Nucleic Acids Res. *46*, 4677–4688. https://doi.org/10.1093/nar/gky264.

26. Maule, G., Arosio, D., and Cereseto, A. (2020). Gene therapy for cystic fibrosis: progress and challenges of genome editing. Int. J. Mol. Sci. *21*, 3903. https://doi.org/10.3390/ijms21113903.

27. Welcome to CFTR2 | CFTR2. https://www.cftr2.org/.

28. Uusi-Mäkelä, M.I.E., Barker, H.R., Bäuerlein, C.A., Häkkinen, T., Nykter, M., and Rämet, M. (2018). Chromatin accessibility is associated with CRISPR-Cas9 efficiency in the zebrafish (Danio rerio). PLoS One *13*, e0196238. https://doi.org/10.1371/journal.pone.0196238.

29. Yarrington, R.M., Verma, S., Schwartz, S., Trautman, J.K., and Carroll, D. (2018). Nucleosomes inhibit target cleavage by CRISPR-Cas9 in vivo. Proc. Natl. Acad. Sci. USA *115*, 9351–9358. https://doi.org/10.1073/pnas.1810062115.

30. Chung, C.-H., Allen, A.G., Sullivan, N.T., Atkins, A., Nonnemacher, M.R., Wigdahl, B., and Dampier, W. (2020). Computational analysis concerning the impact of DNA accessibility on CRISPR-Cas9 cleavage efficiency. Mol. Ther. *28*, 19–28. https://doi.org/10.1016/j.ymthe.2019.10.008.

31. Ciciani, M., Demozzi, M., Pedrazzoli, E., Visentin, E., Pezzè, L., Signorini, L.F., Blanco-Miguez, A., Zolfo, M., Asnicar, F., Casini, A., et al. (2022). Automated identification of sequence-tailored Cas9 proteins using massive metagenomic data. Nat. Commun. *13*, 6474. https://doi.org/10.1038/s41467-022-34213-9.

32. Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Dohmae, N., Ishitani, R., Zhang, F., and Nureki, O. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell *156*, 935–949. https://doi.org/10.1016/j.cell.2014.02.001.

33. Crudele, J.M., and Chamberlain, J.S. (2018). Cas9 immunity creates challenges for CRISPR gene editing therapies. Nat. Commun. *9*, 3497. https://doi.org/10.1038/s41467-018-05843-9.

34. Chew, W.L., Tabebordbar, M., Cheng, J.K.W., Mali, P., Wu, E.Y., Ng, A.H.M., Zhu, K., Wagers, A.J., and Church, G.M. (2016). A multifunctional AAV–CRISPR–Cas9 and its host response. Nat. Methods *13*, 868–874. https://doi.org/10.1038/nmeth.3993.

35. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823. https://doi.org/10.1126/science.1231143.

36. Ensinck, M., De Keersmaecker, L., Heylen, L., Ramalho, A.S., Gijsbers, R., Farré, R., De Boeck, K., Christ, F., Debyser, Z., and Carlon, M.S. (2020). Phenotyping of rare CFTR mutations reveals distinct trafficking and functional defects. Cells *9*, 754. https://doi.org/10.3390/cells9030754.

37. Anders, C., and Jinek, M. (2014). In vitro enzymology of Cas9. Methods Enzymol. *546*, 1–20. https://doi.org/10.1016/B978-0-12-801185-0.00001-5.

38. Brinkman, E.K., Chen, T., Amendola, M., and van Steensel, B. (2014). Easy quantitative assessment of genome editing by sequence trace decomposition. Nucleic Acids Res. *42*, e168. https://doi.org/10.1093/nar/gku936.

39. Clement, K., Rees, H., Canver, M.C., Gehrke, J.M., Farouni, R., Hsu, J.Y., Cole, M.A., Liu, D.R., Joung, J.K., Bauer, D.E., and Pinello, L. (2019). CRISPResso2 provides accurate and rapid genome editing sequence analysis. Nat. Biotechnol. *37*, 224–226. https://doi.org/10.1038/s41587-019-0032-3.

40. Nobles, C.L., Reddy, S., Salas-McKee, J., Liu, X., June, C.H., Melenhorst, J.J., Davis, M.M., Zhao, Y., and Bushman, F.D. (2019). iGUIDE: an improved pipeline for analyzing CRISPR cleavage specificity. Genome Biol. *20*, 14. https://doi.org/10.1186/s13059-019-1625-3.

YMTHE, Volume *31*

Supplemental Information

An optimized SpCas9 high-fidelity variant for direct protein delivery

Eleonora Pedrazzoli, Andrea Bianchi, Alessandro Umbach, Simone Amistadi, Mégane Brusson, Giacomo Frati, Matteo Ciciani, Kalina Aleksandra Badowska, Daniele Arosio, Annarita Miccio, Anna Cereseto, and Antonio Casini

SUPPLEMENTAL MATERIALS

Figure S1

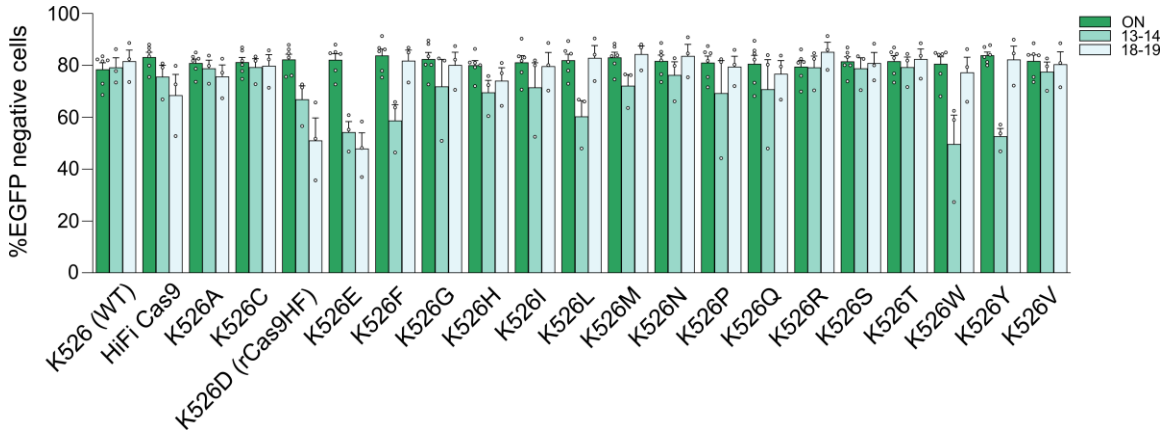

Figure S1. Evaluation of the activity and specificity of SpCas9 K526 mutants. 293multiEGFP cells were transiently transfected with the indicate SpCas9 variants, together with a perfectly matching sgRNA (ON) or two surrogate off-target sgRNAs containing a mismatch couple in position 13-14 or 18-19 of the spacer, counting from the PAM. The percentage of EGFP-negative cells was then measured by cytofluorimetry 7 days post-transfection. Data reported as mean ± SEM for n≥3 biologically independent replicates.

Figure S2

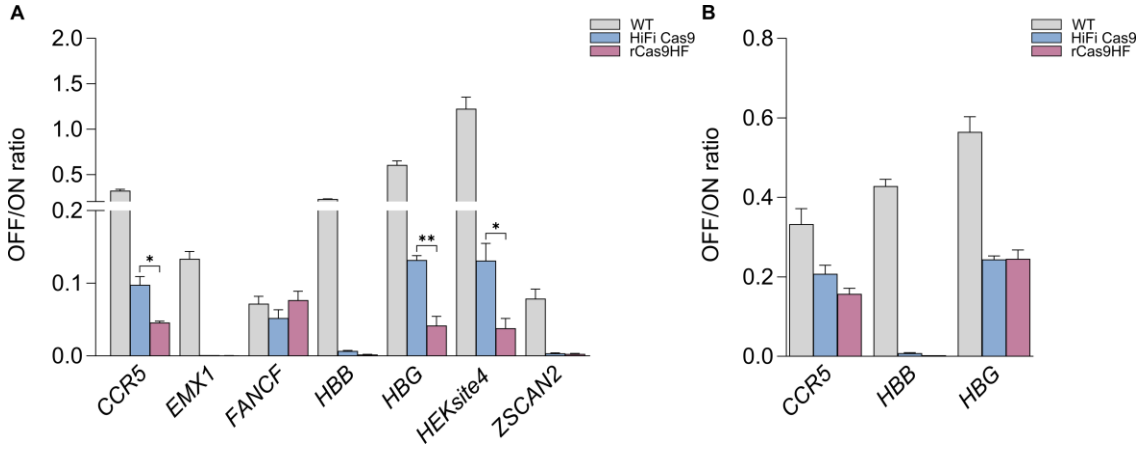

Figure S2. Specificity of rCas9HF on endogenous loci in in U2OS and CD34+ cells compared to wt SpCas9 and HiFi Cas9. (A) Off/on target ratios relative to editing data in **Fig. 2C-D** after electroporation of U2OS cells using wt SpCas9, HiFi Cas9 or rCas9HF RNPs. **(B)** Off/on target ratios relative to **Fig. 3A-B** after CD34+ cells electroporation. Statistical significance was assessed using paired t-test. Data reported as mean ± SEM for n=3 biologically independent replicates.

Figure S3

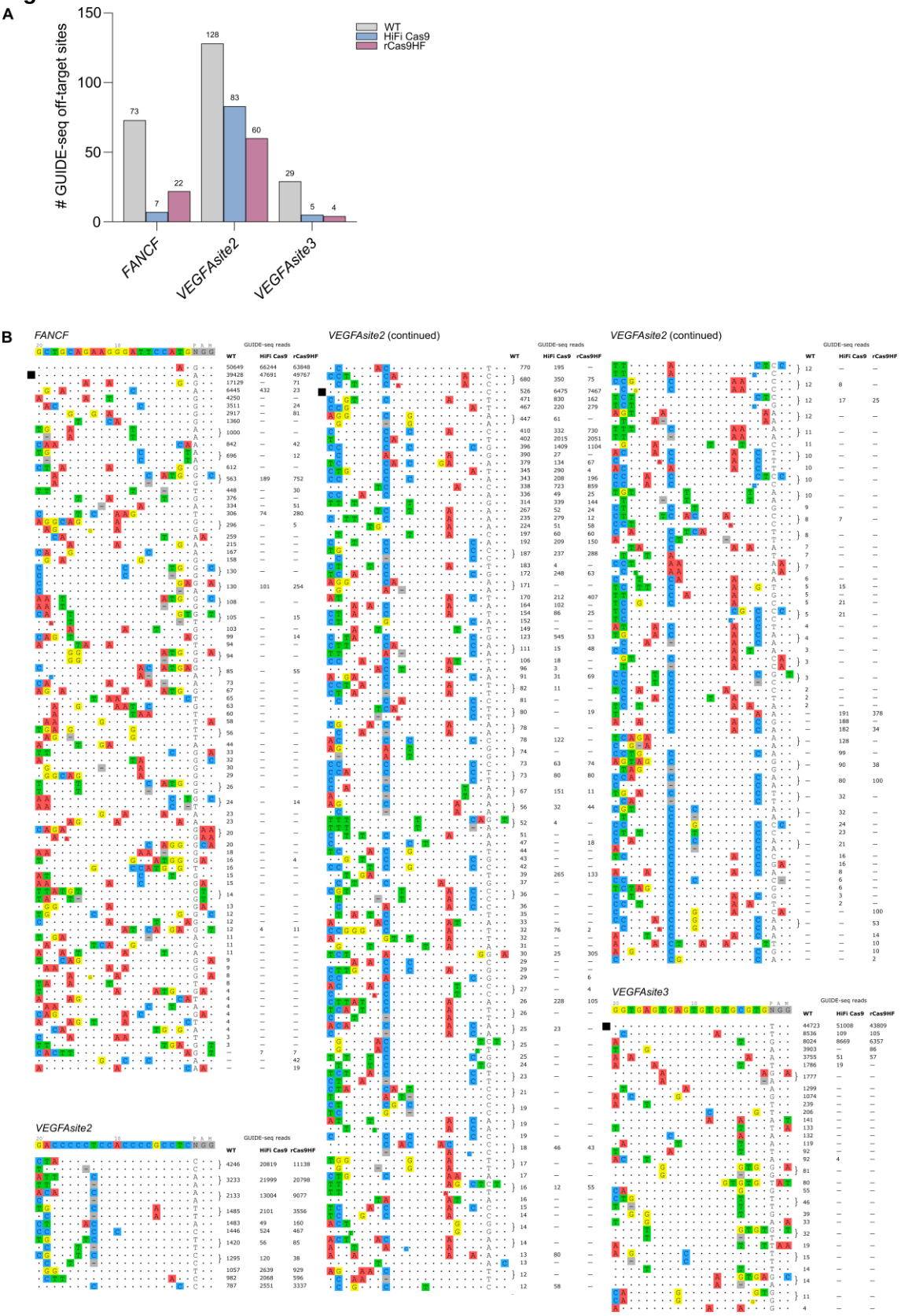
A
GUIDE-seq off-target sites
150
100
50
0
WT
HiFi Cas9
rCas9HF
73
7
22
128
83
60
29
5
4
FANCF
VEGFAsite2
VEGFAsite3
B
FANCF
VEGFAsite2
VEGFAsite2 (continued)
VEGFAsite2 (continued)
VEGFAsite3
GUIDE-seq reads
WT
HiFi Cas9
rCas9HF

Figure S3. GUIDE-seq analysis.

(A) Total number of off-target sites detected by GUIDE-seq upon editing of the FANCF, VEGFAsite2 and VEGFAsite3 target sites using wt SpCas9, HiFi Cas9 or rCas9HF delivered as RNP **(Fig. 2E)**. **(B)** List of off-targets sites detected by GUIDE-seq obtained with the three SpCas9 variants on FANCF, VEGFAsite2 and VEGFAsite3. GUIDE-seq read counts for the different variants are reported on the right of the sequences. (-) indicates that the off-target was not detected in the corresponding sample. Mismatched positions are indicated with colored boxes. Black squares indicate on-target sites. Additional information on the identified off-targets is reported in Tables S3-11.

Figure S4

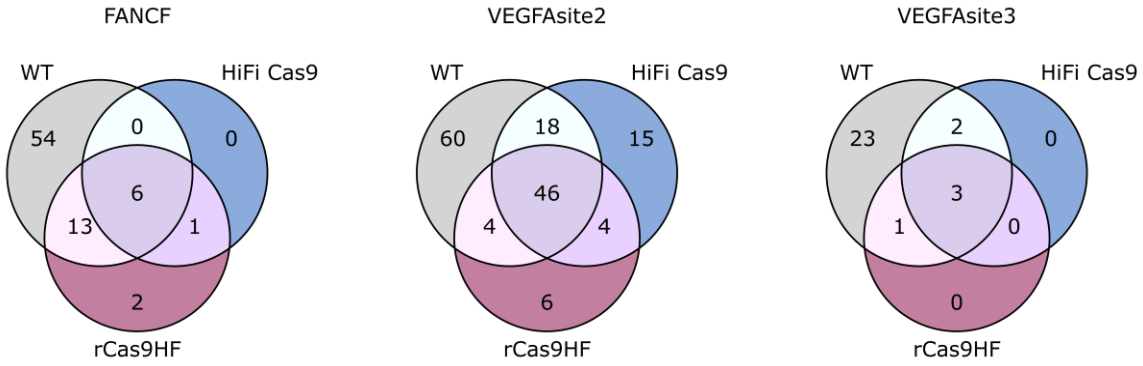

Figure S4. Off-target sites shared by wt SpCas9, rCas9HF and HiFi Cas9 variants.

Venn diagrams showing the distribution of GUIDE-seq identified off-target sites **(Fig. S3)** for wt SpCas9, rCas9HF and HiFi Cas9 variants associated with the sgRNAs targeting the loci indicated above each graph.

Figure S5

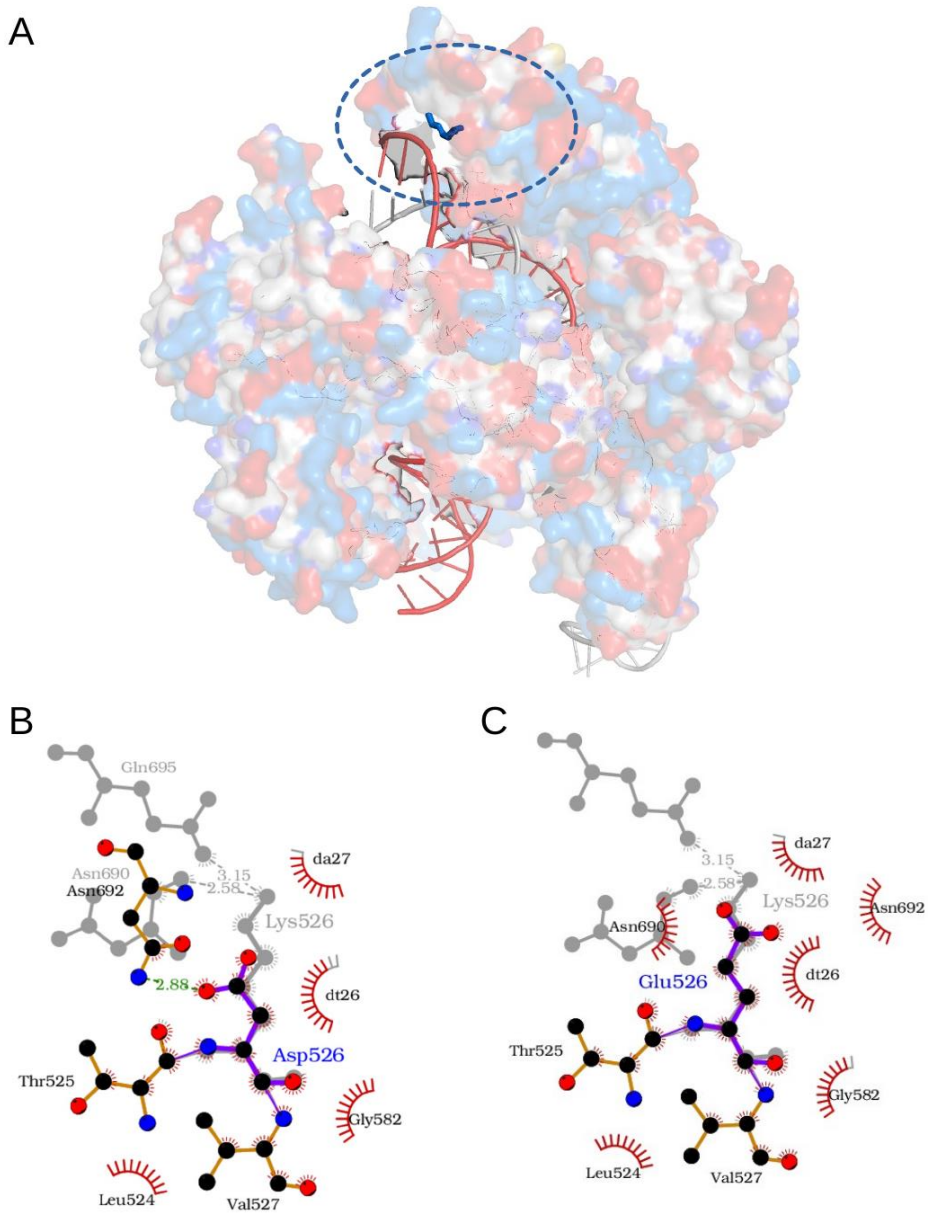

Figure S5. Differences between K526E mutant and rCas9HF at the protein-DNA interface. **(A)** Three-dimensional view of the x-ray structure (PDB: 4UN3) of Cas9 bound to PAM-containing DNA target. The ellipse (blue dashed line) shows the region around the lysine at position 526, Lys-526, represented with bold sticks (in blue). The surface representation of Cas9 shows positively (in red) and negatively (in blue) charged residues. Target DNA is shown as a red ribbon and sgRNA as a white ribbon.

(B-C) 2D diagrams of the interactions predicted by LigPlot+[1] coloured for the mutated structures (Asp or Glu) superimposed on a gray wild-type (Lys). Hydrogen bonds are shown as dashed lines with bond distance (Å) printed in the middle and hydrophobic contacts are represented by arcs with spokes. All interacting (Cas9) residues and (DNA) nucleotides are indicated, the ones involved in hydrogen bonds are drawn with atom details as well as being labelled. Structures of the mutants were obtained in silico with the mutagenesis wizard of the PyMol software. **(B)** Superimposing LigPlot analysis for D526 and K526 (in gray) shows that 2 hydrogen bonds between Lys526 and both Gln695 and Asn690 are replaced by a new hydrogen bond between Asp526 and Asn692. **(C)** Superimposing LigPlot analysis for E526 and K526 (in gray) shows that only hydrophobic contacts are maintained with Asn690 in K526E mutant.

TABLES S1-S11 supplied as Excel files.

Table S1. Sequences of gRNA spacers, oligos for EGFP spacers cloning, ssODNs, oligos for amplification of on-target and relative off-target sites

Table S2. Oligonucleotides used for SpCas9 site-directed mutagenesis and cloning

Table S3. GUIDE-seq raw data WT Cas9 on FANCF

Table S4. GUIDE-seq raw data R691A Cas9 on FANCF

Table S5. GUIDE-seq raw data K526D Cas9 on FANCF

Table S6. GUIDE-seq raw data WT Cas9 on VEGFAsite2

Table S7. GUIDE-seq raw data R691A Cas9 on VEGFAsite2

Table S8. GUIDE-seq raw data K526D Cas9 on VEGFAsite2

Table S9. GUIDE-seq raw data WT Cas9 on VEGFAsite3

Table S10. GUIDE-seq raw data R691A Cas9 on VEGFAsite3

Table S11. GUIDE-seq raw data K526D Cas9 on VEGFAsite3

REFERENCES

1. Laskowski, R.A., and Swindells, M.B. (2011). LigPlot+: Multiple ligand-protein interaction diagrams for drug discovery. J. Chem. Inf. Model. *51*, 2778–2786. 10.1021/ci200227u.

Part II - CoCas9, a compact nuclease from the human microbiome for efficient and precise genome editing

The utilization of CRISPR tools has significantly accelerated the development of therapies for genetic diseases. However, the continuous progress in this field is contingent upon diversifying these tools to address challenges associated with genome editing, particularly in the realm of delivery. The introduction of rCas9HF has emerged as a valuable tool for RNP-based genome editing through high-fidelity endonucleases; nonetheless, certain challenges persist, such as the PAM constraint and the dimensional considerations of SpCas9.

In pursuit of optimizing various aspects of Cas9, with a particular focus on the capability to package the entire system into a single AAV vector, our research endeavors to explore novel avenues. This involves mining natural sources to identify short CRISPR-Cas9 systems that possess enhanced attributes.

Methods

Identification of Cas9 orthologs and tracrRNA prediction from metagenomic data. *cas1*, *cas2* and *cas9* genes were identified from the protein annotation of 154,723 bacterial and archaeal metagenome-assembled genomes (MAGs), reconstructed from the human microbiome[152]. CRISPR arrays were identified using MinCED[153] version 0.4.2 (with default parameters). Only loci having a CRISPR array and *cas1-2-9* genes at a maximum distance of 10 kbp from each other were considered. Loci containing Cas9 proteins shorter than 950 aa were discarded. The resulting 17,173 CRISPR-Cas9 loci were filtered by selecting short proteins (less than 1100 aa) from poorly characterized species. Cas9 proteins from the same species, having similar length but slightly different sequence, were compared by multiple sequence alignment. Proteins presenting deletions in nuclease domains were discarded. The remaining proteins were compared for sequencing coverage and the ortholog with the highest coverage was selected for each species. The Cas9 subtype was predicted using CRISPRCasTyper v1.2.1[154].

Prediction of tracrRNAs for CRISPR-Cas9 loci of interest was performed based on the work by Chyou and Brown[155]. Starting from unique direct repeats in the CRISPR array, BLAST version 2.2.31 (with parameters -task blastn-short -gapopen 2 -gapextend 1 -penalty -1 -reward 1 -evalue 1 -word_size 8) was used to identify anti-repeats within a 3000 bp window flanking the CRISPR-Cas9 locus. RNIE[156] was used to predict Rho-independent transcription terminators (RITs) near anti-repeats. Putative tracrRNA sequences, starting with an anti-repeat and ending with either a RIT (when found) or a poly-T, were combined with direct repeats to form sgRNA scaffolds. The secondary structure of sgRNA scaffolds was predicted using RNAsubopt version 2.4.14 (with parameters --noLP -e 5)[157]. sgRNAs lacking the functional modules identified by Briner *et al.*[158] namely the repeat:anti-repeat duplex, nexus and 3' hairpin-like folds, were discarded.

The identified Cas9 orthologs coding sequences were modified by the addition of an SV5 tag at the N-terminus and two nuclear localization signals (1 at the N-term and 1 at the C-term) and were human codon-optimized (**Table 1**). The constructs for the coding sequences, as well as the sgRNA scaffolds, were obtained as synthetic fragments from either Genscript or Genewiz.

Construction of the randomized PAM library. The randomized PAM library was prepared as described in Kleinstiver et al.[136] Briefly: one synthetic DNA oligonucleotide containing an EcoRI site and a 8-nt randomized sequence (top oligo) was obtained from Eurofins, together with another DNA oligo that anneals to the 3' region flanking the randomized sequence

leaving an SphI-compatible end (bottom oligo). The bottom strand of the annealed oligo duplex was filled-in by Klenow(exo-) incubation and digested with EcoRI for ligation into a SphI/EcoRI-cut p11-lacY-wtx1 vector (Addgene Plasmid #69056). The ligation product was then electroporated into MegaX DH10B T1R Electrocomp™ Cells (Thermo Fisher Scientific) to reach a theoretical library coverage of 100X. Colonies were harvested and the plasmid DNA was purified by maxi-prep (Macherey-Nagel). Two PCR steps (Phusion HF DNA polymerase - Thermo Fisher Scientific) were performed to prepare the plasmid PAM library for NGS analysis: the first, using a set of forward primers and two different reverse primers, to amplify the region containing the protospacer and the randomized PAM and the second to attach the Illumina Nextera DNA indexes and adapters (**Table 2**). PCR products were purified using Agencourt AMPure beads in a 1:0.8 ratio. The library was analyzed with a 150-bp single read sequencing, using a v3 flow cell on an Illumina MiSeq sequencer.

***In vitro* assay for PAM identification.** The *in vitro* PAM evaluation of the novel Cas9 orthologs was performed according to the protocol from Karvelis *et al*.[159] with few modifications. In brief: the synthetic DNA encoding the human codon optimized version of the Cas9 genes was obtained from Genscript and cloned into an expression vector for *in vitro* transcription and translation (IVT) (pT7-N-His-GST Thermo Fisher Scientific). Reaction was performed according to the manufacturer protocol (1-Step Human High-Yield Mini IVT Kit - Thermo Fisher Scientific). The Cas9-gRNA RNP complex was assembled by combining 20 µL of the supernatant containing soluble Cas9 protein with 1µL of RiboLock RNase Inhibitor (Thermo Fisher Scientific) and 2µg of guide RNA. The RNP was used to digest 1 µg of the randomized PAM plasmid DNA library for 1 hour at 37°C.

A double stranded DNA adapter[159] (**Table 2**) was ligated to the DNA ends generated by the targeted Cas9 cleavage and the final ligation product was purified using a GeneJet PCR Purification Kit (Thermo Fisher Scientific).

One round of a two-step PCR (Phusion HF DNA polymerase - Thermo Fisher Scientific) was performed to enrich the sequences that were cut using a set of forward primers annealing on the adapter and a reverse primer designed on the plasmid backbone downstream of the PAM (**Table 2**). A second round of PCR was performed to attach the Illumina indexes and adapters. PCR products were purified using Agencourt AMPure beads in a 1:0.8 ratio. The library was analyzed with a 71-bp single read sequencing, using a flow cell v2 micro, on an Illumina MiSeq sequencer.

PAM sequences were extracted from Illumina MiSeq reads and used to generate PAM sequence logos, using Logomaker version 0.8. PAM heatmaps[160,161] were used to display

PAM enrichment, computed dividing the frequency of PAM sequences in the cleaved library by the frequency of the same sequences in a control uncleaved library.

Estimation of PAM frequency and editable adenines. The number of targetable genomic sites was computed counting the number of occurrences of PAMs in the hg38 human reference genome for each Cas9 nuclease. For base editors, mutations in the ClinVar database[162] were filtered to select G>A and C>T single nucleotide variants (SNVs) annotated as pathogenic or likely pathogenic, for a total of 40,871 SNVs. Editing windows of base editors were estimated as positions showing at least 10% mean editing on all tested guides (**Table 3**). Selected SNVs were considered targetable if the surrounding sequence contained a PAM, on the appropriate strand, that would place the mutated adenine inside the estimated editing window.

CoCas9 cleavage pattern assay. *In vitro* CoCas9 cleavage products were analyzed by PCR and Sanger sequencing. EGFP and B2M PCR products were obtained amplifying pEGFP-N1 and HEK293T DNA respectively (HotFire polymerase, Solis BioDyne), using primers indicated in **Table 4**. CoCas9 protein was produced using the 1-Step Human High-Yield Mini IVT Kit (Thermo Fisher Scientific), while *EGFP* and *B2M* sgRNAs were *in vitro* transcribed (HighYield T7 RNA Synthesis Kit, Jena Bioscience) starting from the amplification (Phusion HF DNA polymerase - Thermo Fisher Scientific) of the plasmid bearing CoCas9 sgRNA (primers reported in **Table 4**). The RNP complex was assembled using 300nM sgRNA, 3.8 µl of *in vitro* translated CoCas9 and RiboLock RNase Inhibitor 1 U/µl (Thermo Fisher Scientific) in 10X nuclease reaction buffer (200 mM HEPES, 1 M NaCl, 50 mM $MgCl_2$, 1 mM EDTA). After 10 minutes of incubation, 30 nM of corresponding PCR was added and the reaction (30 µl) was left for 1 hour at 37°C; 1 µl of RNase A/T1 Mix (Thermo Fisher Scientific) and 0.7 µg/µl of Proteinase K (Thermo Fisher Scientific) were finally added. The PCR fragments were loaded on agarose gel, purified and sent for Sanger sequencing (Eurofins) using primers used for amplification.

Mammalian expression plasmids and constructs. A pX330-derived plasmid was used to express the Cas9 orthologs and relative sgRNA in mammalian cells. Briefly, pX330 was modified by substituting the SpCas9 and its sgRNA scaffold with the human codon-optimized sequence of the orthologs and its sgRNA scaffold (either full length or trimmed). The Cas9s coding sequences and the sgRNA scaffolds were obtained as synthetic fragments from either Genscript or Genewiz; Cas9s were fused with a V5 tag at the N-terminus and two nuclear localization signals (1 at the N-term and 1 at the C-term). Spacer sequences were cloned into the pX-Cas9 plasmids as annealed DNA oligonucleotides containing a variable 20 or 24-nt spacer sequence using double BsaI sites present in the plasmid. The list of

spacer sequences used in the EGFP disruption assay and in the endogenous loci targeting is reported in **Table 5**.

For base editor constructs, the nCoCas9 (D23A) obtained by PCR mutagenesis was fused with the ABE8e domain from CP1041-ABE8e (Addgene #138493) and introduced in a pCMV-derived plasmid, while the optimized trimmed version of CoCas9 sgRNA (CoCas9 TS-opt) was cloned in a pUC19-derived plasmid (**Table 6**).

pAAV-v1 and pAAV-v2 (AAV-v1 and AAV-v2 plasmids) were synthesized by Vectorbuilder; gRNA1_CoCas9_HEKsite2 (shortened H2-g1) was selected for pAAV-v1 and pAAV-v2 after transfection experiments. pAAV-v3 and pAAV-v4 (AAV-v3 and AAV-v4 plasmids) were obtained by adding the U6-gRNA3.1_CoCas9_RHO (**Table 5**) cassette and substituting EGFP with the CoCas9 sequence in pTIGEM[163].

A plentiCRISPR v1 plasmid (Addgene Plasmid 49535) was used to deliver CoCas9 in human primary cell lines. Briefly, plentiCRISPR v1 was modified by substituting the SpCas9 and its sgRNA scaffold with the human codon-optimized sequence of CoCas9 and its sgRNA scaffold. Spacer sequences were cloned in the plentiCRISPR v1 using BsmBI restriction sites.

Cell lines and culture conditions. HEK293T cells (obtained from ATCC), U2OS.EGFP cells (a kind gift of Claudio Mussolino, University of Freiburg), harboring a single integrated copy of an EGFP reporter gene, AAVpro-293T (obtained from Takara), HEK293-RHO-EGFP cells and HSF (GM05659, from the Coriell Institute) were cultured in DMEM (Life Technologies) supplemented with 10% FBS (Life Technologies), 2 mM GlutaMax (Life Technologies) and penicillin/streptomycin (Life Technologies).

HBE (BE121 from the Italian Cystic Fibrosis Research Foundation) were cultured in LHC9/RPMI 1640 medium (1:1) without serum as previously described[164,165], supplemented with rho-associated protein kinase 1 inhibitor (Y-27632, 5 μM) (Merck) and SMAD-signalling inhibitors, bone morphogenetic protein antagonist (DMH-1, 1 μM), and transforming growth factor β antagonist (A 83-01, 1 μM) to promote basal cell proliferation[166].

HEK293-RHO-EGFP cells were obtained by stable transfection of HEK293 cells with a pcDNA5/TO-RHO-EGFP reporter plasmid, obtained by cloning a fragment of the *RHO* gene up to exon 2 (retaining introns 1 and 2) fused to part of RHO cDNA containing exons 3-5 in frame with the EGFP coding sequence into a pCDNA5/TO expression plasmid. Cells were pool-selected with 5 μg/ml Hygromycin (Invivogen) and single clones were subsequently

isolated and expanded. All cells were incubated at 37°C and 5% CO2 in a humidified atmosphere. All cells tested mycoplasma negative (PlasmoTest, Invivogen).

Cell line transfections. For EGFP disruption assay 200,000 U2OS. EGFP cells were nucleofected with 1 µg of pX-Cas9 plasmids bearing a sgRNA designed to target EGFP using the 4D-Nucleofector™ X Kit (Lonza), DN100 program, according to the manufacturer's protocol. After electroporation, cells were seeded in 96-well plates and transferred to 24-well plates after 48 hours for growth expansion.

For editing analyses of genomic loci, 100,000 HEK293T cells were seeded in a 24-well plate 24 hours before transfection. Cells were then transfected either with 1 µg of pX-Cas-sgRNA plasmid expressing both the Cas9s and the sgRNAs, or with 500 ng of a pX-Cas9 plasmids and 250 ng of a pUC-sgRNA constructs, using the TransIT-LT1 reagent (Mirus Bio) according to the manufacturer's protocol. Cell pellets were collected 3 days post-transfection for indel evaluations. For base editing experiments, cells were co-transfected as described above with 750 ng of pCMV-ABE8e containing the specific Cas9s and 250 ng of pUC-sgRNA.

Lentiviral vectors production and cell transduction. Lentiviral particles were produced in HEK293T cells at 80% confluency in p150 plates. 25 µg of transfer vector (lentiCRISPR v1) plasmid, 7.5 µg of VSV-G and 18.8 µg of Δ8.91 packaging plasmid were transfected using PEI. The culture medium was replaced the day after with complete DMEM. The viral supernatant was collected after 48 h and filtered through a 0.45 µm PES filter. Lentiviral vectors were concentrated and purified with 20% sucrose cushion by ultracentrifugation for 2 h at 4 °C and 150,000 × g. Pellets were resuspended in OptiMEM and aliquots stored at −80 °C. Vector titres were measured as reverse transcriptase units (RTU) by SG-PERT method[167].

For transduction experiments, HSF and HBE were seeded (60,000 cells/well and 100,000 cells/well, respectively) in 12-well plates and the day after were transduced with 2 RTU of lentiviral vectors. 72 h later, cells were cultured with puromycin (2 µg/ml) and collected 10 days from transduction for editing analysis.

AAV production and cell transduction. For AAV-DJ, AAVpro-293T cells (10^7 cells/dish in three P150 dishes) were PEI transfected with pHelper, pAAV Rep-Cap and the proper pAAV-v1-4 construct. Three days post-transfection, media and cells were collected, centrifuged and processed separately. Cells were washed and lysed with an acidic citrate buffer (55 mM citric acid, 55 mM sodium citrate, 800 mM NaCl, pH 4.2)[168]. The lysates were

cleared by centrifugation, treated with HEPES buffer 1M, DNaseI and RNaseA (Thermo Fisher Scientific) and then mixed with the collected medium and 500 mM NaCl. AAVs were precipitated with 8% polyethylene glycol (PEG) 8000 overnight at 4°C, then collected by centrifugation and resuspended in TNE Buffer (100 mM Tris·Cl, pH 8.0, 150 mM NaCl, 20 mM EDTA) followed by 1:1 chloroform extraction. AAV titration was performed by qPCR following the Addgene protocol[169] (https://www.addgene.org/protocols/aav-titration-qpcr-using-sybr-green-technology/).

AAV8 vectors were produced by InnovaVector S.R.L. as previously described[170]. The titer (genome copies/ml) of each viral preparation was determined by averaging the titer achieved by dot-blot analysis and by qPCR quantification using TaqMan (Applied Biosystems, California, USA)[170].

For AAV-DJ transduction in cell lines 10^5 cells were transduced in a 24-well plate. After 3 days the medium was changed to remove the vector and cells were collected 6 days post-transduction for editing analysis.

Animal model. Mice were housed at TIGEM animal house (Pozzuoli, Italy). The knock-in mouse model used in this study was kindly provided by Theodore G. Wensel[171]. This mouse model harbors the full-length human rhodopsin (h*RHO*) gene including the P23H mutation (hRHO-P23H). hRHO-P23H C-terminus is fused to the red fluorescent protein (RFP) coding sequence (hRHO-P23H-RFP). hRHO-P23H-RFP−/− mice were maintained by crossing homozygous females and males. Experimental heterozygous hRHO-P23H-RFP+/− animals were obtained by crossing homozygous hRHO-P23H-RFP−/− with C57BL/6J mice obtained from Envigo Italy SRL (Udine, Italy). Genotype analysis was performed using a previously described protocol[171].

Subretinal AAV injections in mice. Subretinal injections with AAV8 (AAV-v3) were performed on the temporal side of the eye using 1 µl of vector-containing solution via a trans-scleral, trans choroidal approach[172] in accordance with both the Association for Research in Vision and Ophthalmology Statement for the Use of Animals in Ophthalmic and Vision and with the Italian Ministry of Health regulation for animal procedures (Ministry of Health authorization number: 252/2022-PR). The surgical intervention was performed under general anesthesia; mice received an intraperitoneal injection of ketamine (10 mg/Kg) combined with medetomidine (1 mg/Kg). hRHO-P23H-RFP+/− mice were injected sub-retinally at around 5-weeks of age with a combination of two AAV8 vectors: one expressing CoCas9 (AAV-v3, 2×10^9 genome copies/eye) and the other the enhanced green fluorescent protein (EGFP, 2×10^8 genome copies/eye, to track the injected area) both under

the control of the ubiquitous CMV promoter. Control eyes received the AAV8 CMV-EGFP vector alone. Right and left eyes were randomly assigned to each treatment group. One month post-injection mice were euthanized, the neural retina of each eye was collected and separated into EGFP+ and EGFP- portions for indel formation analyses.

Evaluation of editing efficacy. For EGFP disruption assay, EGFP fluorescence was measured four days after nucleofection using a BD FACSCanto (BD) flow cytometer. DNA was extracted using the QuickExtract DNA Extraction Solution (Lucigen) for cell culture 3 days after transfection, 6 days after AAV transduction and 10 days after lentiviral transduction. DNeasy Blood & Tissue kit (Qiagen) was used to isolate genomic DNA from cells 10 days after transduction, and from mouse retina samples at 30 days from AAV transduction according to the manufacturer's instructions. To amplify the target loci, PCR reactions were performed using the HOT FIREPol DNA polymerase (Solis BioDyne), using the oligonucleotides listed in **Table 4**. The amplified products were purified, Sanger sequenced (Microsynth) and analyzed with the TIDE web tool (http://shinyapps.datacurators.nl/tide/) to quantify indels or with the EditR web tool (http://baseeditr.com) to quantify base editing events.

In vivo editing efficiency of CoCas9 at the h*RHO* gene target in the mice retinas was analyzed by deep sequencing. To amplify the target locus, a first round of PCR reactions was performed using HOT FIREPol DNA polymerase (Solis BioDyne), using 100 ng of DNA as template and the oligonucleotides listed in **Table 2**. The amplified products were purified (Agencourt AMPure beads), quantified and used for a second round of PCR (Phusion HF DNA polymerase - Thermo Fisher Scientific) to attach the Illumina indexes. After final purification and quantification, a pool of each sample with the same nM was analyzed with a 150-bp single read sequencing, using a v2 flow cell on an Illumina MiSeq sequencer.

Off-target analysis. GUIDE-seq experiments were performed as previously described[151]. Briefly, 2×10^5 HEK293T cells were transfected using Lipofectamine 3000 transfection reagent (Invitrogen) with 1 µg of all-in-one pX plasmid, expressing CoCas9 and sgRNA, and 10 pmol of dsODNs; scramble sgRNA was used as negative control. The day after transfection, cells were detached and selected with 1 µg/ml puromycin as described in Casini *et al*.[124] Three days after transfection, cells were collected, and genomic DNA extracted using NucleoSpin Tissue Kit (Macherey-Nagel) following manufacturer's instructions. Using Covaris S200 sonicator, genomic DNA was sheared to an average length of 500 bp. End-repair reaction was performed using NEBNext Ultra End Repair/dA Tailing Module and adaptor ligation using NEBNext® Ultra™ Ligation Module, as described by Nobles *et al*.[173] Amplification steps were then performed following the GUIDE-seq original protocol[151].

Visualization of aligned off-target sites relative to GUIDE-seq analysis showed in **Figure 11f-g** is available as a color-coded sequence grid in the Appendix (**Figure S1**).

Western Blot. Cell pellets from cells transfected with Cas9s expression plasmids and control sgRNAs were collected after 24, 48 and 72 hours and lysed with RIPA buffer. 10 µg of the whole lysates quantified by BCA assay (Thermo Fisher Scientific) were run on a SDS-PAGE gel and transferred to a PVDF membrane. Membranes were probed with mouse anti-V5 antibodies (Thermo Fisher Scientific, #46-0705) recognizing the V5 tag fused to Cas9 and anti-α-Tubulin antibodies (Sigma Aldrich, #T6074). Detection was performed with Pierce ECL reagent (Thermo Fisher Scientific) using the Uvitec Alliance Instrument.

Subcellular fractionation. HEK293T cells (500,000 cells/well) were seeded on a 6-well plate and transfected with 2.5 µg of Cas9s expression plasmids (pX-Cas9s) and 1.25 µg of pUC-sgRNA. Pellets were collected after 72 hours and resuspended in Nuclear Resuspension Buffer (NSB) (10 mM HEPES pH 8, 10 mM KCl, 1.5 mM MgCl2, 0.34 M Sucrose, 10% Glycerol, 1 mM DTT, 0.1% TritonX, 1X Protease inhibitors), followed by cytoplasmic and nuclear fractions separation by sequential centrifugations. Samples' concentration was assessed through Bradford assay (Sigma-Aldrich) and 7.5 µg of each fraction were analyzed by Western blot. Cas9s were detected with anti-V5 antibodies (Thermo Fisher, #46-0705), using anti-GAPDH (Santa Cruz Biotechnology, #sc-32233) and anti-H3 (Abcam, #ab1791) as loading control and to verify the purity of the subcellular fractions. Goat anti-Mouse (KPL, #0741809) or anti-Rabbit (Santa Cruz Biotechnology, #sc-2004) HRP-conjugated were used as secondary antibodies.

Statistical analysis. Statistical significance tests were performed using GraphPad Prism (version 9.4.1). Data in **Figures 10h** and **13b** were analyzed with a one-way ANOVA followed by a two-sided Holm-Šídák test. Data in **Figures 11b** and **13a** were analyzed using two-sided t-tests corrected for multiple comparisons using the Holm-Šídák method, while a two-sided t-test was used in **Figures 11c and 16g**. For all analyses, adjusted p-values less than 0.05 were considered statistically significant.

Results

Identification of CRISPR-Cas9 systems from bacterial metagenomes

We[2] searched for compact Cas9s (<1100 aa) in a large database constituted by 154,723 microbial genomes[152], identifying 436 Cas9 proteins between 950 aa and 1100 aa, characterized by a complete CRISPR locus (*cas1*, *cas2*, tracrRNA and an adjacent CRISPR array). The shortest 11 Cas9s were selected, being characterized by a large CRISPR array (>5 spacers) and a readily identifiable tracrRNA. The PAM sequences for all the nucleases were determined by exploiting an *in vitro* PAM assay[159] (see **Methods**), where the resulting logos showed a remarkable variety among the recognized sequences (**Fig. 9a**). Once obtained the PAM, we were able to assess the editing activity in mammalian cells through an EGFP disruption assay in a U2OS-EGFP cell line, comparing our new orthologs with SpCas9. Six out of eleven Cas9s did not show any editing efficiency, four presented a range from 5 to 50%, and the most significant result (approximately 80% of EGFP-negative cells) was observed using a Cas9 a 1004 aa long protein from the uncharacterized bacterium *Collinsella* sp. AM34-10, which we named CoCas9 (**Fig. 9b**).

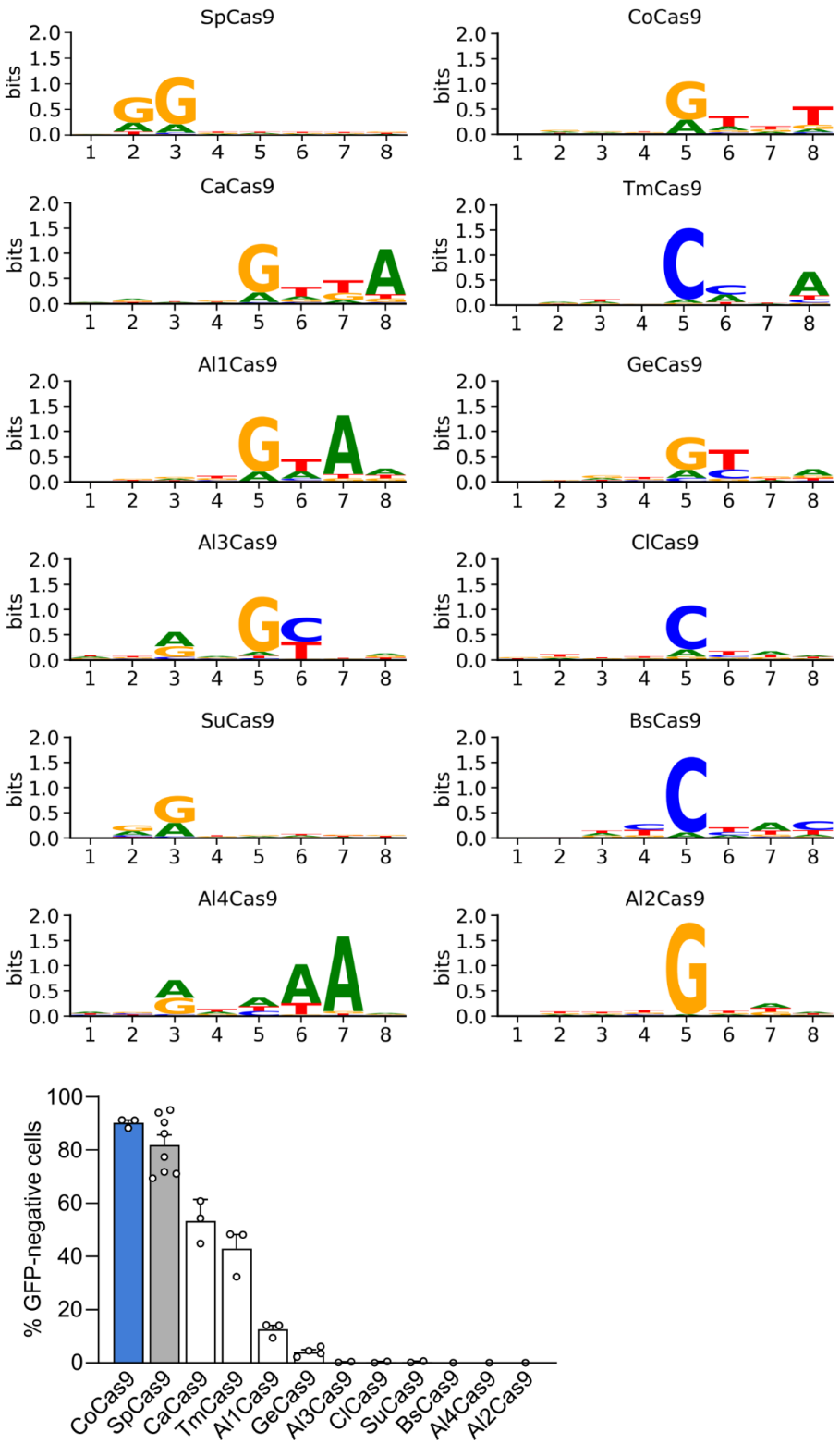

Figure 9. PAM logos of selected Cas9 orthologs identified in the microbiome database by Pasolli et al[152]. *(**a**) PAM sequences identification and derived logos were obtained as described in Methods; SpCas9 PAM sequence was added to the analysis as positive control. (**b**) Activity of 10 newly identified Cas9 orthologs and the benchmark nuclease SpCas9 compared through an EGFP disruption assay in U2OS-EGFP cells; data reported as mean ± SEM for n≥3 independent experiments.*

Characterization of the CoCas9-based tools

CoCas9 shows the highest resemblance with type II-C Cas9s found in *Campylobacter jejuni*[139] and *Neisseria meningitidis*[143] (**Fig. 10a**). Despite this, it only shares a modest amino acid sequence similarity of 30.9% and 31.5% with the two proteins, respectively. The CoCas9 CRISPR locus is structured as shown in **Figure 10b**. The CRISPR array is composed of 24 spacer-direct repeat units and it is located downstream of the *cas9*, while the tracrRNA is found between the array and the *cas1-cas2* genes. The tracrRNA has a predicted secondary structure similar to other type II tracrRNAs[158] (**Fig. 10c** and **Table 6**). From the PAM detection assay we obtained a frequency heatmap of the 256 four-nucleotide combinations contained in the PAM of CoCas9, revealing three main consensus sequences: 5'-N_4GWNT, 5'-N_4GCDT, and 5'-N_4ATDT (**Fig. 10d**). Through the sequencing of *in vitro* cleaved fragments (**Fig. 10e**), we demonstrated that CoCas9 generates blunt-end products, cutting the target DNA 3 bp upstream of the PAM (**Fig. 10f**), similarly to SpCas9 cleavage products. To optimize the cleaving conditions, we also modified the structure of the gRNA, at the level of scaffold and spacer (**Fig. 10g-h**). The CoCas9 original dual gRNA (**Fig. 10c**) was combined in the full scaffold version (FS), which was obtained by the fusion of the repeat and anti-repeat through a GAAA tetraloop (**Table 6**). The FS was compared with three alternative designs, one containing base substitutions, aimed at increasing the stability of the secondary structure (FS-opt), the second being a trimmed version characterized by a shorter repeat:anti-repeat loop (TS), and the third corresponds to a stabilized version of TS (TS-opt) (**Table 6**). When tested, these sgRNA variants did not significantly change the editing activity (**Fig. 10g**). To investigate which spacer length would be the best for cleavage efficiency, we tested 22, 23, and 24 nt-long spacers on two genomic loci. Even if on *FAS* there was no significant difference among the spacer sizes, on *HBB* there was an improvement in CoCas9 activity when using a 23 nt spacer (**Fig. 10h**). Based on these results, we decided to proceed using a 23 nt spacer (G+23 nt when needed) with a TS-opt sgRNA.

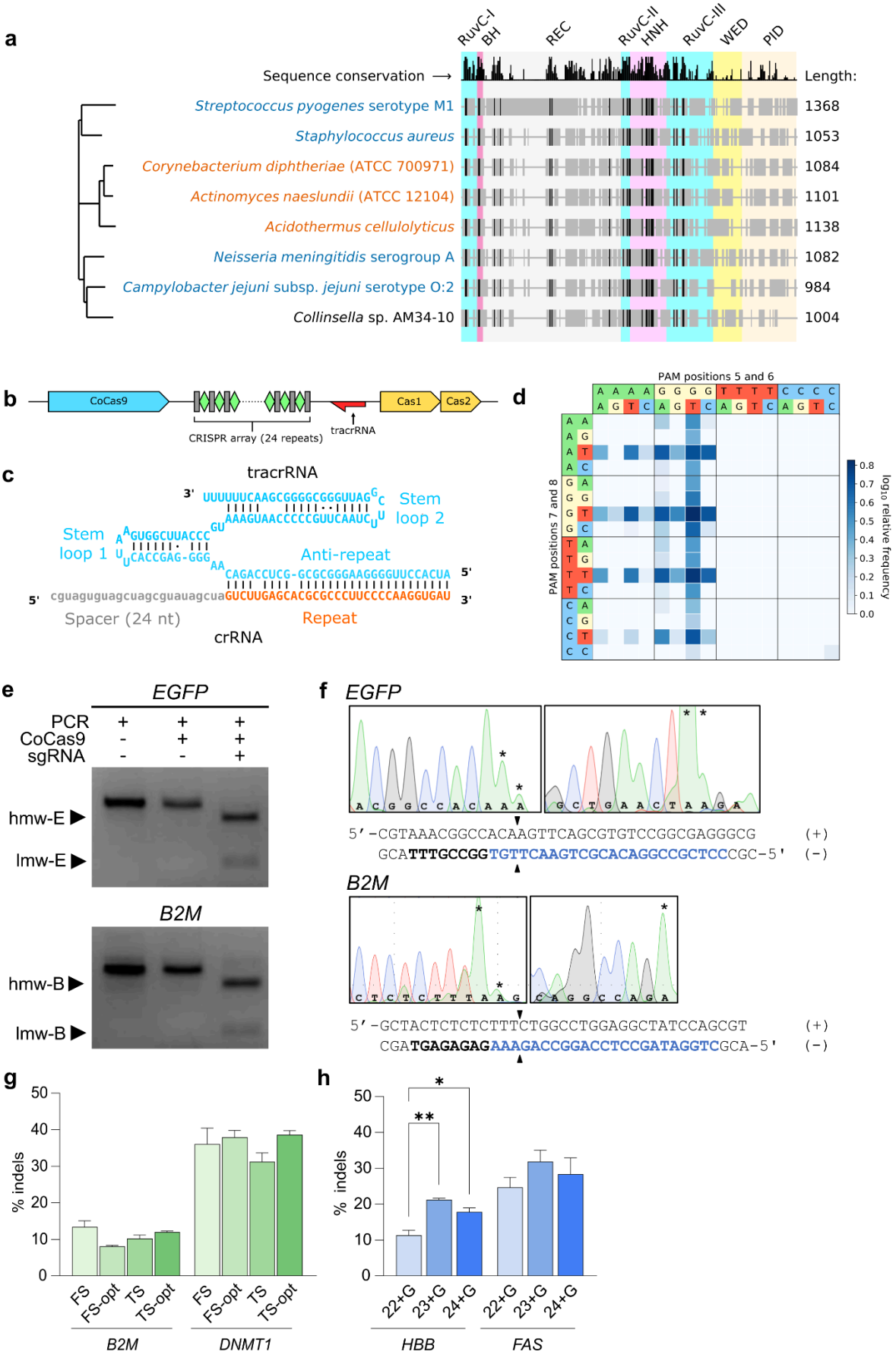
a
RuvC-I
BH
REC
RuvC-II
HNH
RuvC-III
WED
PID
Sequence conservation
Length:
Streptococcus pyogenes serotype M1 1368
Staphylococcus aureus 1053
Corynebacterium diphtheriae (ATCC 700971) 1084
Actinomyces naeslundii (ATCC 12104) 1101
Acidothermus cellulolyticus 1138
Neisseria meningitidis serogroup A 1082
Campylobacter jejuni subsp. jejuni serotype O:2 984
Collinsella sp. AM34-10 1004
b
CoCas9
Cas1
Cas2
CRISPR array (24 repeats)
tracrRNA
c
tracrRNA
Stem loop 2
Stem loop 1
Anti-repeat
Spacer (24 nt)
Repeat
crRNA
d
PAM positions 5 and 6
PAM positions 7 and 8
log10 relative frequency
e
EGFP
PCR
CoCas9
sgRNA
hmw-E
lmw-E
B2M
hmw-B
lmw-B
f
EGFP
5′-CGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCG (+)
GCATTTGCCGGTGTTCAAGTCGCACAGGCCGCTCCCGC-5′ (-)
B2M
5′-GCTACTCTCTCTTTCTGGCCTGGAGGCTATCCAGCGT (+)
CGATGAGAGAGAAAGACCGGACCTCCGATAGGTCGCA-5′ (-)
g
% indels
FS
FS-opt
TS
TS-opt
B2M
DNMT1
h
% indels
22+G
23+G
24+G
HBB
FAS

Figure 10. CoCas9 characterization: a functionally active compact nuclease from the Collinsella genus. *(**a**) Phylogenetic tree of selected Cas9 subfamilies mostly used for genome editing applications (blue) or with experimentally determined three-dimensional structure (orange). Protein alignments: gray aligned protein sequences, black conserved sequences, colored conserved domains. Length: number of aa. (**b**) Scheme of the CoCas9 CRISPR locus. (**c**) crRNA and tracrRNA predicted secondary structure. (**d**) Log-scale heatmap showing relative frequency enrichment of the 256 PAM nucleotide combinations at the four most informative positions compared to the non-cut PAM library. (**e**) Agarose gels showing EGFP and B2M PCR fragments obtained from the in vitro cleavage. (**f**) The cleavage products (hmw-E, lmw-E, hmw-B, and lmw-B from panel e) were Sanger sequenced. Left panels: chromatogram of the hmw (plus strand); right panels: chromatogram of the lmw (negative strand). Asterisks indicate adenines (A) are byproducts of Sanger sequencing. In the depiction of the cut position, black triangles represent the cut site, the protospacer sequence is in blue and the PAM is in bold. Sequencing primers in Table 4. (**g**) Side-by-side comparison of alternative CoCas9 sgRNA scaffolds. The editing activity was tested in two endogenous genomic loci (B2M and DNMT1). (**h**) The optimal sgRNA spacer length for CoCas9 was assessed by targeting HBB and FAS genes in HEK293T cells using spacers ranging from 22 to 24 nt. Data plotted as mean ± SEM for n=3 biological replicates. Statistical significance was assessed with a one-way ANOVA followed by a two-sided Holm-Šídák test; **P<0.01, *P<0.05, no significant comparisons not shown.*

Encouraged by the promising editing results, we moved forward with the evaluation of CoCas9 activity at the level of the genome. The efficacy of CoCas9 was measured against a panel of 26 loci, showing up to 55% indels at specific targets (*HEKsite1* and *IL2RG*) and variable levels of activity (**Fig. 11a**). However, to have a thorough comparison with the widely used SpCas9, we designed 24 pairs of overlapping spacers in an equal number of loci, so that the 3' portion of the CoCas9 and SpCas9 spacers were perfectly matching (**Fig. 11b**). Although CoCas9 exhibited a significantly lower mean editing efficacy overall (mean difference = 12.2%, **Fig. 11c**), comparable percentages of indels were detected at the majority of the loci (n=14, **Fig. 11b**). The expression levels of these two orthologs were verified to exclude editing variability depending on protein abundance (**Fig. 11d-e**). More precisely, **Figure 11d** illustrates the assessment of CoCas9 and SpCas9 protein levels over time (24 - 48 - 72 hours), revealing no substantial differences between the two. Additionally, we examined the expression of Cas9s in both nucleic and cytoplasmic fractions, observing a predominant presence of CoCas9 in the cytoplasm compared to SpCas9 (**Fig. 11e**). We observed a lower band during the western blot analysis of CoCas9 (**Fig. 11d-e**) and SpCas9 (**Fig. 11e**), which may result from protein degradation.

To test the precision of CoCas9, we performed a genome-wide comparative off-target analysis with SpCas9 through GUIDE-seq[151]. To this aim, we selected four loci (*HPRT*, *VEGFAsite2*, *ZSCAN2,* and *Chr6*) where both nucleases showed similar editing efficacy with overlapping spacer sequences (**Fig. 11b**).

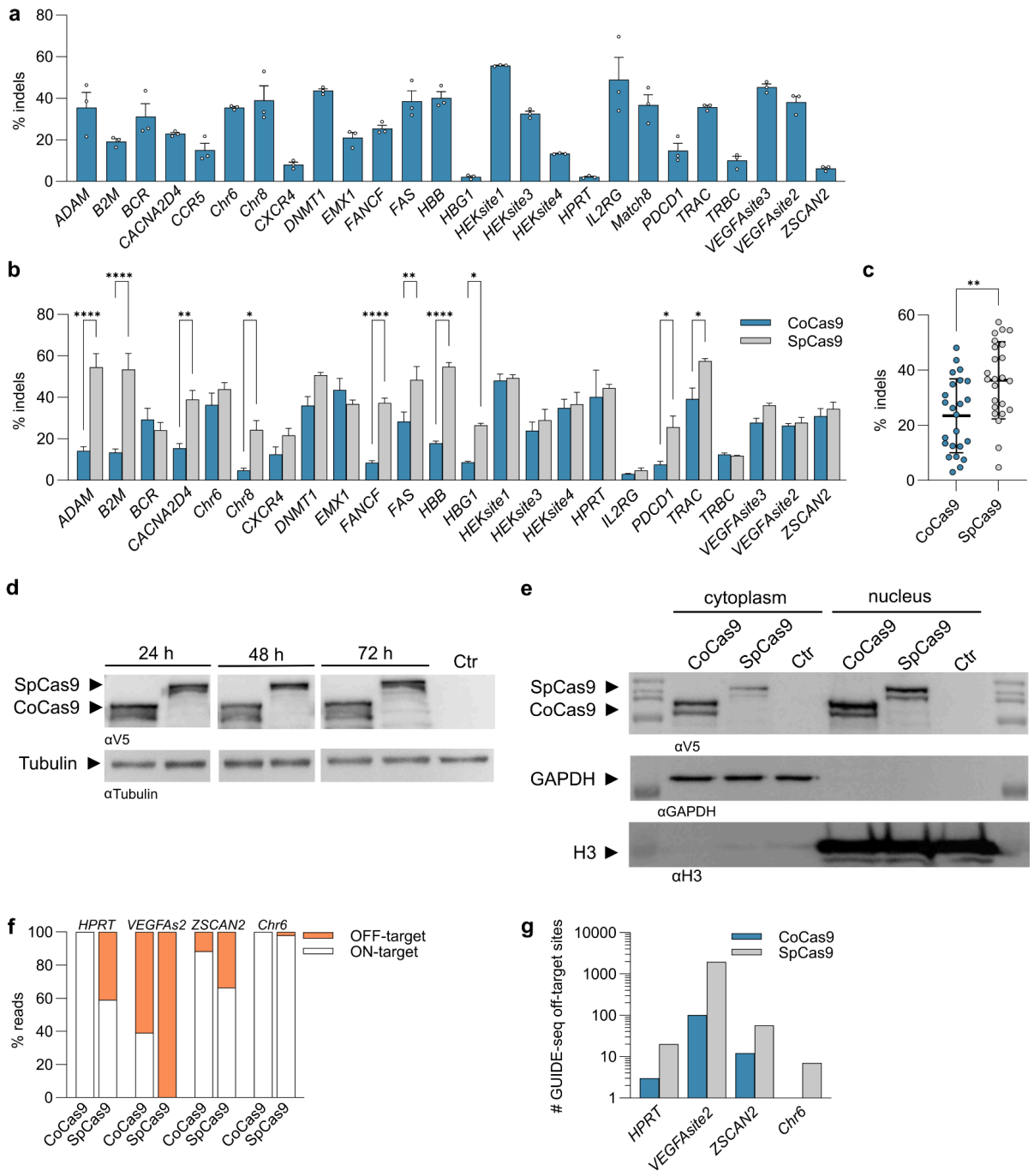

Figure 11. Comparison of CoCas9 and SpCas9 activity and precision. *(**a**) CoCas9 percentages of indels in a panel of 26 genomic loci in HEK293T cells. (**b**) Side-by-side comparison of the editing activity of CoCas9 and SpCas9 on a panel of 24 genomic loci in HEK293T cells using overlapping spacers. (**c**) Dot plot summary of the editing efficiency of CoCas9 and SpCas9 at the 24 loci from panel (b). (**d**) Western blot analysis from HEK293T cells transfected with CoCas9 and SpCas9 plasmids at three different time points (see Methods). Cas9 proteins were detected with anti-V5 antibodies; anti-tubulin was used as loading control. (**e**) Western blot analysis from cytoplasmic and nuclear fractions (see Methods) after transfection with CoCas9 and SpCas9 plasmids. Cas9 proteins were detected with anti-V5 antibodies; anti-GAPDH or anti-H3 were used as controls of cytoplasmic or nuclear fraction respectively. Ctr=non-transfected cells. (**f**) CoCas9 and*

*SpCas9 on- versus off-target reads derived from GUIDE-seq in four genomic loci. (**g**) Total number of off-target sites obtained with CoCas9 and SpCas9 from GUIDE-seq analysis (panel f). Data reported as mean ± SEM for n≥3 independent experiments. In panel (b), the statistical significance was assessed using two-sided t-tests corrected for multiple comparisons using the Holm-Šídák method. In panel (c) the statistical significance was assessed using a two-sided t-test. ****P<0.0001, ***P<0.001, **P<0.01, *P<0.05, not significant comparisons not shown.*

In all examined loci, CoCas9 produced considerably fewer off-target cleavages than SpCas9 (**Fig. 11f-g**). The superior performance of CoCas9 was particularly striking at the off-target benchmark *VEGFAsite2*[124,151], where CoCas9 showed a 10-fold reduction of off-target sites compared to SpCas9 (**Fig. 11g**).

After investigating the nuclease activity, we wanted to test CoCas9 as a base editor. The CoCas9 nickase (nCoCas9) was generated by mutating the D23 residue of the RuvC-I domain, and it was fused with an adenosine deaminase[97,108] to generate CoABE8e. In the 19 tested loci, we detected up to 55% of A>G transition efficiency depending on the target locus (**Fig. 12a-b**). Similar to the comparison between CoCas9 and SpCas9, we tested overlapping guides on 6 targets with CoABE8e and SpCas9-derived base editor[108] (SpABE8e). Although CoABE8e showed overall lower base-editing efficiency, CoABE8e can match SpABE8e A>G conversion in the *HEKsite1* target (**Fig. 12b**). Moreover, CoCas9 exhibited a different window of activity at specific sites, since the CoCas9 spacer is longer than SpCas9's, 24 nt and 20 nt respectively (*HEKsite4* and *TRAC* targets in **Fig. 12b**).

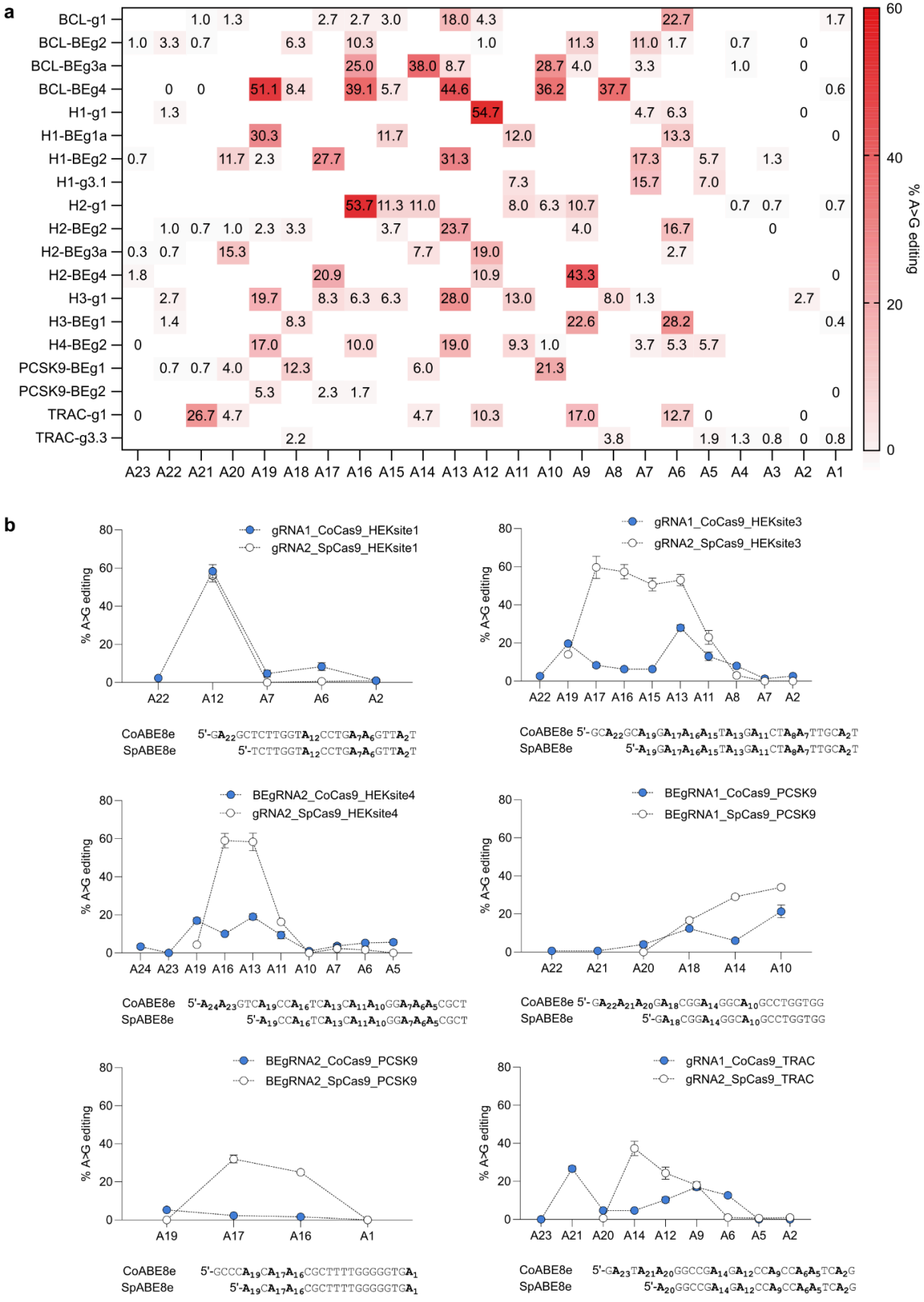

Figure 12. CoABE8e characterization and comparison with SpABE8e. *(**a**) Summary of the frequency of A>G transitions by CoABE8e in HEK293T cells with 19 sgRNAs targeting a*

*panel of endogenous loci. Data reported as the mean of n=3 biologically independent experiments. (**b**) Base editing efficiency of CoABE8e and SpABE8e using overlapping spacers at the indicated loci. CoABE8e sgRNAs from panel (a) are: H1-g1, H3-g1, H4-BEg2, PCSK9-BEg1, PCSK9-BEg2, and TRAC-g1. The A residues are counted starting from the PAM proximal side of the spacer going in the 5' direction. Data reported as mean ± SEM for n=3 biological replicates.*

CoCas9 was set side by side with Cas9 orthologs similar in size. The most characterized Cas9 with similar size are SaCas9 (PAM = 5'-N_2GRRT)[137] and Nme2Cas9 (PAM = 5'-N_4CC)[144], with 1053 and 1082 aa length, respectively. These orthologs were compared with CoCas9 by analyzing 11 genomic loci targeted by sgRNAs with overlapping spacers (**Fig. 13a**). Nme2Cas9 had overall significantly lower activity than the other two nucleases, while CoCas9 generated more indels than SaCas9 in 7 loci out of 11 (**Fig. 13a**). To deepen the comparison with small Cas9s (SaCas9, Nme2Cas9, and CjCas9, CjCas9 PAM = N_3VRYAC[139]) we included target sites with non-overlapping PAM and spacers (n=18 sgRNAs for each ortholog) (**Table 5**). Overall, CoCas9 was significantly more active than Nme2Cas9 and CjCas9 and comparable to SaCas9 (**Fig. 13b**). However, as shown in **Figure 13c**, CoCas9 is theoretically able to target a higher number of sequences in the genome, as its PAM is more frequent compared to SaCas9's (11.8% and 2.9%, respectively) (see **Methods**). Also, the specificity profile of CoCas9 was discovered to be significantly better, since we compared CoCas9 and SaCas9 through GUIDE-seq using 5 sgRNAs targeting different loci (**Fig. 13d-e**). CoCas9 generated far fewer off-targets than SaCas9 for two guides, more precisely 31 and 0 off-targets for SaCas9 and CoCas9, respectively, for the *PDCD1* gRNAs, while the gRNAs on *VEGFAsite3* presented 15 off-target sites for SaCas9 against 3 for CoCas9. Nearly no off-targets were observed for both nucleases in the remaining loci (**Fig. 13e**).

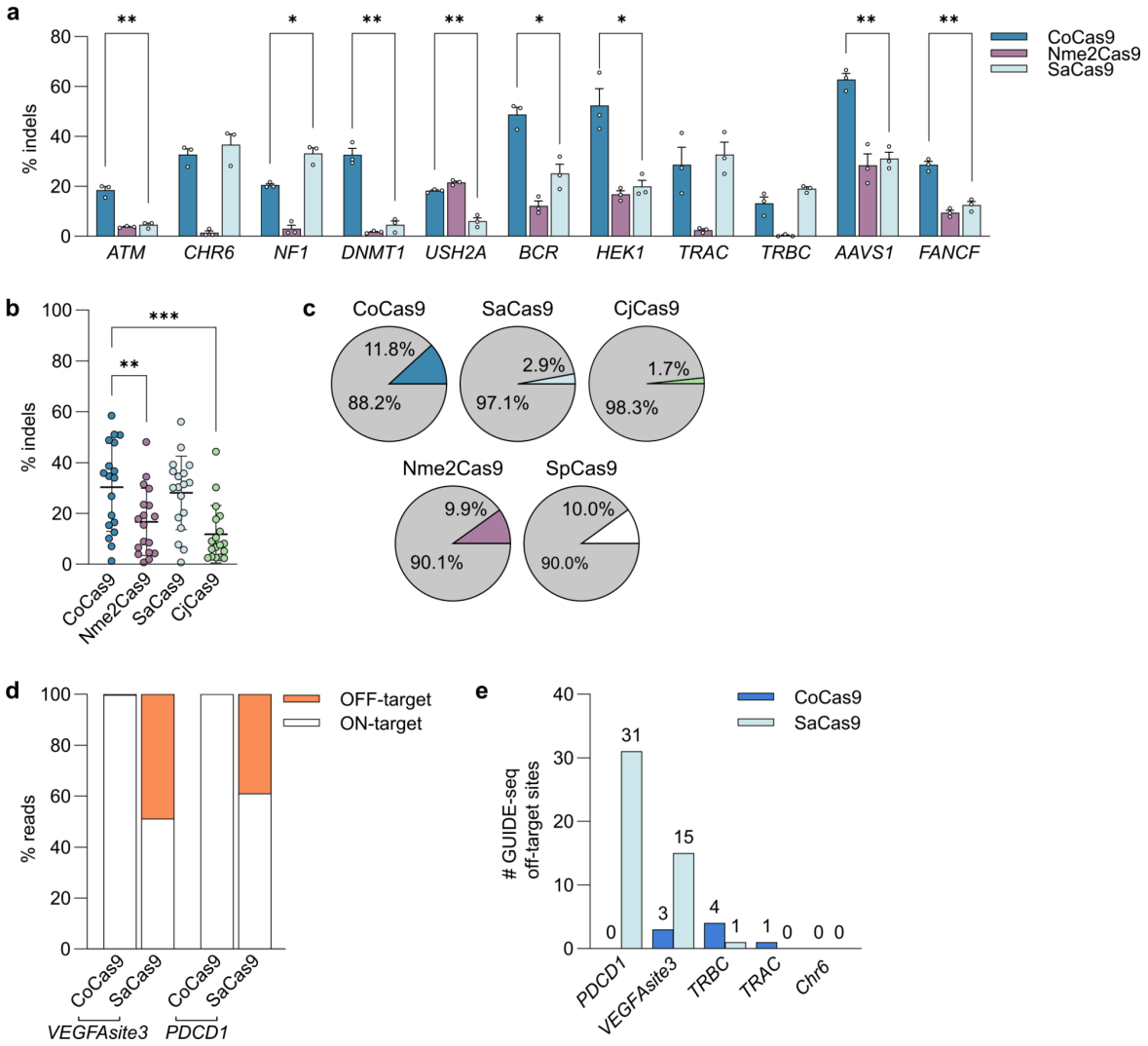

Figure 13. Comparison of nuclease activity among CoCas9 and other compact Cas9s. *(**a**) Editing efficacy (% indels) of CoCas9, Nme2Cas9, and SaCas9 overlapping sgRNAs. (**b**) Dot plot of nuclease efficiency of CoCas9, Nme2Cas9, SaCas9, and CjCas9 using 18 non-overlapping sgRNAs on 6 different genomic loci. (**c**) Percentages of sites in the human genome targetable by the indicated compact Cas9s and SpCas9. (**d**) Comparison of CoCas9 and SaCas9 percentages of on- versus off-target reads obtained by GUIDE-seq at PDCD1 and VEGFAsite3 loci. On- versus off-target reads were not shown for TRBC, TRAC, and Chr6. (**e**) Total number of GUIDE-seq-detected off-target sites for CoCas9 and SaCas9 on 5 loci. Statistical significance was assessed using a two-sided t-test corrected using the Holm-Šídák method for each locus in panel (a), while it was evaluated with a one-way ANOVA followed by a two-sided Holm-Šídák test in panel (b); ***P<0.001, **P<0.01, *P<0.05, not significant comparisons not shown.*

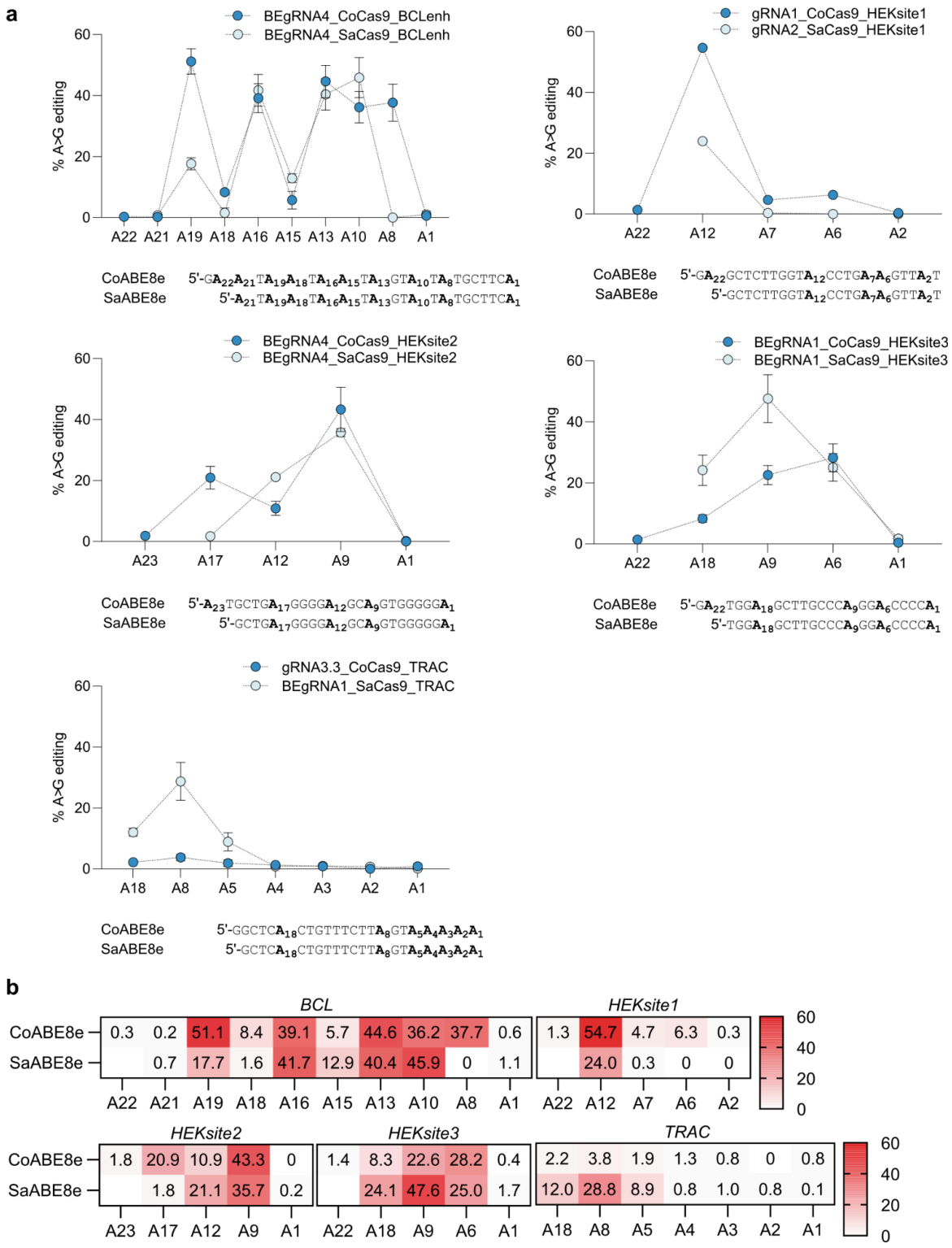

Figure 14. CoABE8e activity compared with the compact base editor SaABE8e. *(**a**) Base editing efficiency of CoABE8e and SaABE8e at the indicated loci using overlapping sgRNAs. CoABE8e sgRNAs from Fig. 12a are: BCL-BEg4, H1-g3, H2-BEg4, H3-BEg1, TRAC-g3.3. The A residues are counted starting from the PAM proximal side of the spacer going in the 5' direction. Data reported as mean ± SEM for n=3 biological replicates. (**b**) Heatmaps summarizing CoABE8e and SaABE8e base editing activity in panel (a).*

To compare CoCas9 and SaCas9 also as base editors, we tested CoABE8e and SaABE8e[108] having sgRNA with overlapping spacers on a panel of 5 loci (**Fig. 12a** and **Table 5**). The two base editors showed a similar editing profile, with CoABE8e more active at 2 out of 5 sites (*BCL* and *HEKsite1*), while almost inactive at the *TRAC* locus (**Fig. 14a-b**).

Aiming to better define the positioning of CoABE8e within the context of compact base editors, we used non-overlapping spacers, allowing us to cross-test CoABE8e with Nme2ABE8e, SauriABE8e[146], and CjABE8e[174]. We obtained very heterogeneous editing profiles from the analysis of 9 sgRNAs targeting 3 genomic loci for each base editor (**Fig. 15a**). The evaluation of the highest A>G conversion produced by each ortholog in the 3 examined loci showed overall similar activity among the analyzed base editors (**Fig. 15b**). To estimate the therapeutic potential of CoABE8e, we retrieved 40,871 G>A pathogenic mutations from ClinVar[162] and predicted the fraction that can be targeted by all tested base editors (see **Methods**). Notably, 10.2% of the considered mutations can be targeted solely by CoABE8e (**Fig. 15c**).

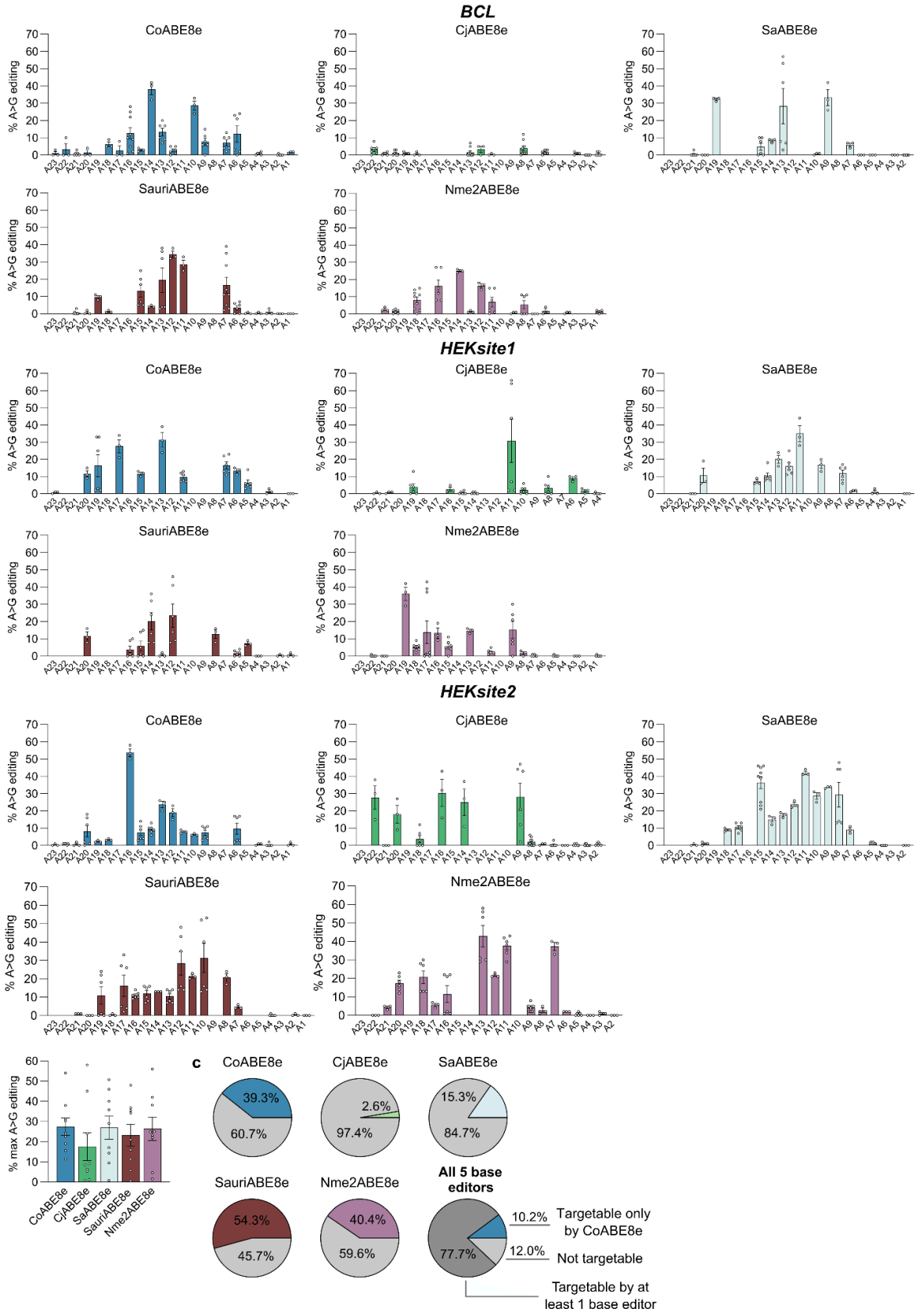

Figure 15. Base editing activity window and efficiency of compact ABE8es. *(a) Base editing activity window for CoABE8e, CjABE8e, SaABE8e, SauriBAE8e and Nme2ABE8e at BCL, HEKsite1 and HEKsite2; non-overlapping sgRNAs were used (9 total sgRNAs for each*

*ABE, see Table 2). The A residues are counted starting from the PAM proximal side of the spacer going in the 5' direction. Values are represented as points, error bars represent mean ± SEM of n=3 biological replicates for each site, with each position representing 1–3 genomic sites. (**b**) Summary of the highest A>G conversion for each sgRNA in (a). (**c**) Editable fraction of human G>A pathogenic mutations by the indicated compact ABE8es, either alone or in combination (All 5 base editors) (see Methods).*

CRISPR-CoCas9 in relevant cellular and animal models

CoCas9 exhibits a compelling profile both as a compact nuclease, coupled with a high on/off-target ratio. Thus, the next phase involved testing its efficacy in primary human cells, specifically human bronchial epithelial cells (HBE) and human skin fibroblasts (HSF) (**Fig. 16a-b**). Lentiviral vectors were employed to transduce HSFs and HBEs with CoCas9 and its sgRNAs targeting six genomic loci. In HBEs, CoCas9 achieved a minimum editing efficiency of 80% (**Fig. 16a**), whereas in HSFs the highest efficacy was observed at the *IL2RG* locus (90%), and the lowest at the *TRAC* locus (40%) (**Fig. 16b**).

The promising properties of CoCas9-based tools led us to test the delivery of CoABE8e through an all-in-one AAV (AAV-v1 and -v2, schematized in **Fig. 16c**). AAV transduction of CoABE8e with the sgRNA targeting *HEKsite2* showed high percentages of A>G transitions, with AAV-v1 reaching 68% of base editing (**Fig. 16e**), thus significantly improving the activity obtained *via* plasmid transfection (53.7%) (**Fig. 12a**).

As mentioned above, for efficient *in vivo* genome editing, the utilization of compact Cas9 for an all-in-one AAV system is crucial. CoCas9 emerged as a suitable choice due to its compact size, so we tested AAV-CoCas9 constructs tailored for editing within the *RHO* gene. The *RHO* gene represents an interesting therapeutic target since it is mutated in a common form of autosomal dominant retinitis pigmentosa[175–177] and our aim for this step was the *in vivo* editing of the mouse retina. We tested two different configurations, AAV-v3 and AAV-v4 (**Fig. 16d**), reaching up to 30% of editing efficiency in HEK293 cells stably expressing a human *RHO* (h*RHO)*-EGFP minigene (**Fig. 16f**). To assess the *in vivo* editing activity of CoCas9, we exploited the design of AAV-v3 (**Fig. 16d** and **g**) to deliver CoCas9 in a knock-in mouse model harboring a single copy of the h*RHO* gene (see **Methods**). We administered an all-in-one AAV8 with CoCas9 targeting h*RHO* to 5 weeks-old mice via subretinal injection (**Fig. 16g**). Four weeks after the administration we observed up to 35.6% of editing efficiency at the h*RHO* target, with a mean of 10.5% in 7 retinas (**Fig. 16h**).

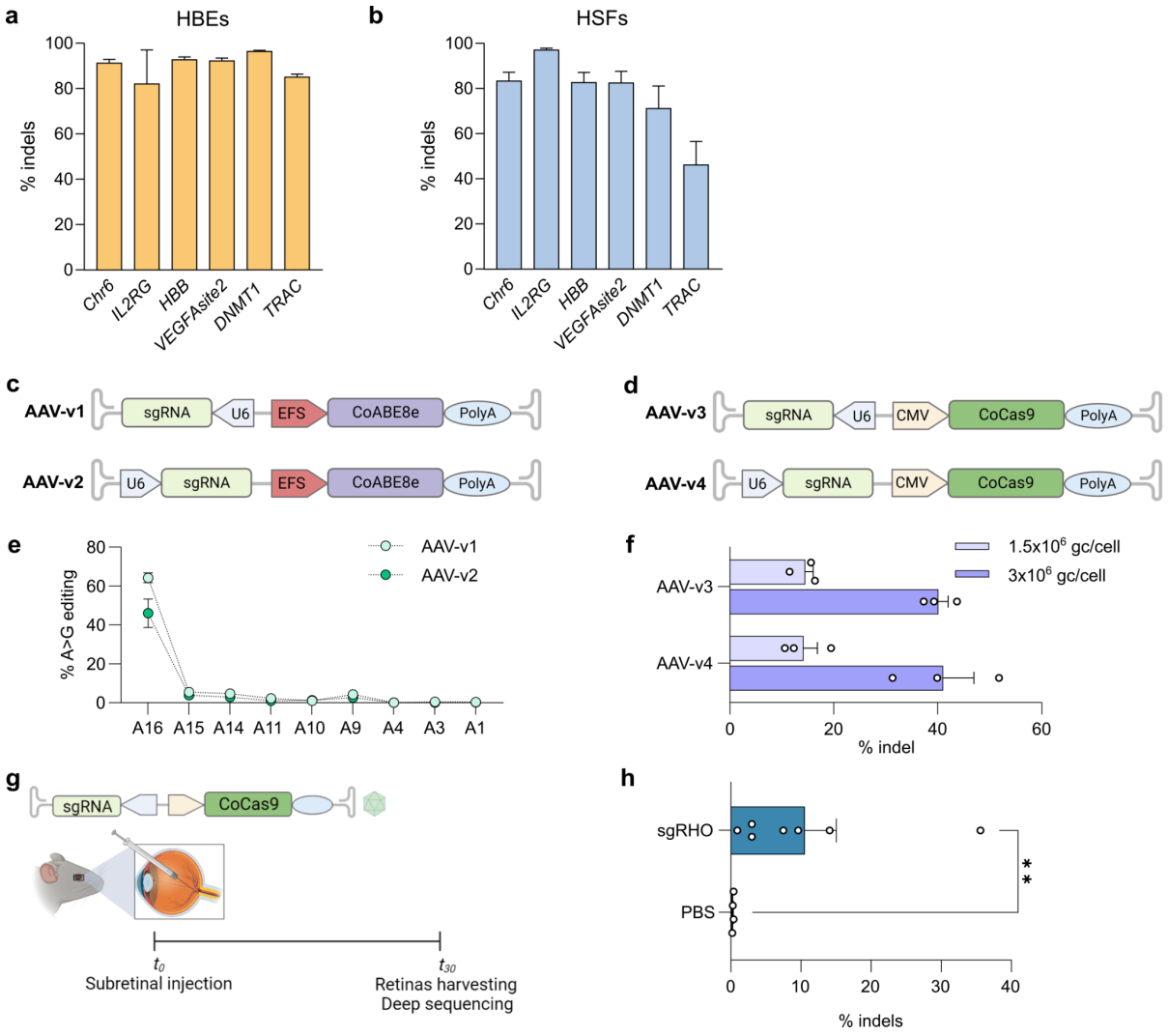

Figure 16. CoCas9 editing activity in human primary cells and mouse retina. *(**a**) Editing efficiency (% indels) of CoCas9 on 6 genomic loci in human bronchial epithelial (HBEs) and (**b**) human skin fibroblasts (HSFs). (**c**) Schematic depiction of two designs of the all-in-one AAV vector expressing CoABE8e under the EFS promoter and the gRNA1_CoABE8e_HEKsite2 under the U6 promoter. (**d**) Scheme of two designs of the all-in-one AAV vector expressing CoCas9 under the CMV promoter and gRNA3.1_CoCas9_RHO under the U6 promoter. The AAV-v3 design is also reported in panel (g). (**e**) Frequency of A>G transitions at the HEKsite2 locus in HEK293T cells transduced with AAV-v1 and AAV-v2 (AAV-DJ serotype) (7x10^5 genome copies/cell, gc/cell). The CoABE8e sgRNA from Fig. 12a is H2-g1. (**f**) Percentages of indels at the RHO locus generated by AAV-v3 and AAV-v4 (AAV-DJ serotype) transduction in HEK293-RHO-EGFP cells with two amounts of gc/cell. (**g**) Flowchart of AAV8-CoCas9 delivery in murine retinas. (**h**) Editing efficiency of AAV8-CoCas9 injected in the mouse retina. Statistical significance was assessed using a two-sided t-test in panel (h); ****P<0.0001, ***P<0.001, **P<0.01, *P<0.05, not significant comparisons not shown. In panels (a, b, e, f, and h) data are shown as mean ± SEM; n=3 biological replicates in panels (a, b, e, and f), n=7 (sgRHO) and n=4 (PBS) independent retinas in panel (h).*

Discussion

The remarkable advances in genome editing have paved the way for a new era of scientific exploration and therapeutic potential, revolutionizing our ability to precisely manipulate genetic information. CRISPR-SpCas9 was the first CRISPR system to be engineered to become a genome editor, and, due to its user-friendly nature and simplicity, it remains the most extensively studied and widely adopted tool in laboratories worldwide. However, clinical applications present obstacles that the nuclease SpCas9 alone is not able to overcome, given the necessity of extreme precision when performing genome modifications in this context.

The large variety of disease-causing mutations and delivery challenges demand enlarging the toolbox of sophisticated and reliable genome editors. In this thesis work, I focused on two aspects: implementing high-fidelity Cas9s suitable for RNP delivery and exploring databases of microbial genomes to discover novel CRISPR-Cas9 systems.

Advancements in CRISPR-Cas9 precision include high-fidelity variants and reduced intracellular Cas9 levels. RNP electroporation represents a valuable transfection method, since the pre-assembly of RNP complexes *in vitro* addresses various challenges. First, the U6 promoter-driven expression of sgRNAs may exhibit variability due to potential premature termination caused by Pol III when sgRNAs contain U-rich sequences[178,179]. Moreover, the use of RNA polymerase for sgRNA expression necessitates the inclusion of mismatched or supernumerary G at the 5′-end of sgRNAs[43,44,180]. Nevertheless, the best advantage of RNP delivery is its quick degradation kinetics when compared to mRNA or plasmid transfection[133], a feature that confers rapid cell clearance and lower off-target effects thanks to a transient peak of Cas9 concentration in the cell[181]. RNP-delivered wt Cas9s show high editing levels and gain an advantage in terms of precision, but it is challenging for high-fidelity variants, since the cleavage activity was reported to decrease[147]. To date, only HiFi Cas9[147] has been identified to efficiently edit the genome through RNP electroporation[147], however, given the variable cleavage efficiency within the human genome, there is a pressing need to diversify the toolkit with additional high-fidelity Cas9 variants with similar attributes.

To address this, we[1] initiated our approach based on SpCas9 mutations identified during evoCas9 development[124], aiming to select amino acid substitutions that confer high editing efficiency to SpCas9 variants. The K526D substitution demonstrated the ability to maintain both high editing efficiency and specificity. This can be attributed to the mutation occurring within the REC3 domain, which interacts with the RNA:DNA duplex[124,182]. Additionally, the shorter side chain of aspartate is instrumental in preserving catalytic activity, since it fits more effectively within the surrounding residues. Encouraged by rCas9HF editing profile, we

tested it in CD34+ cells, since Cas9 RNP is mainly used as a delivery method for *ex vivo* treatments of beta-thalassemia and sickle cell disease[183,184]. Our results illustrate the heterogeneous editing and specificity exhibited by the rCas9HF and HiFi Cas9 variants. This diversity could hold significance at specific therapeutic sites, emphasizing the need for a meticulous assessment of the SpCas9 variant to be employed at each distinct locus. Therefore, analyses like those presented by Schmid-Burgk *et al.* and Kim *et al.*[148,149] are helpful. In these works, they compared the specificity and activity of the most common high-fidelity SpCas9s, through experimental[148] and a deep-learning model[149], respectively. In the future, conducting analogous comparative studies would be particularly compelling in a scientific landscape where the array of available alternatives is expected to expand further[185]. High-fidelity variants, particularly through RNP delivery, are not only useful for editing via indels but also through HDR using DNA donor templates. Indeed, we showed that rCas9HF and HiFi Cas9 delivered as RNP provide an advantage over the wt SpCas9, by reducing the likelihood of re-editing at the repaired locus[186]. This advantage stems from their lower tolerance to mismatches between the sgRNA and the integrated donor template, which carries silenced mutations.

Beyond precise SpCas9 mutants, there is a compelling need within the CRISPR community for robust alternative Cas-based tools. SpCas9 is surely renowned for its remarkable editing efficiency, however, its reliance on a 5'-NGG PAM poses limitations on the range of targetable sequences. SpCas9 is also prone to off-target effects, and its large molecular size hinders effective delivery, especially when considering challenges in compatibility with all-in-one AAV vectors. Various strategies have been adopted to address these challenges, among which direct evolution and rational protein engineering provided several effective high-fidelity Cas9 variants[119,120] and proteins with broader PAM requirements[105,125,185]. However, direct evolution is restricted by the balance of inherent characteristics of the protein and the screened features[187], and rational engineering requires precise structure-activity correlation studies and relies on preconceived assumptions about induced changes. Moreover, the SpCas9 PI domain can be altered to relax the PAM profile, but some downsides must be taken into consideration, such as self-targeting in DNA constructs and unpredictable edits in eukaryotes[188]. Additionally, thorough genome scanning increases target identification times and off-target risks[188]. Regarding the SpCas9 size obstacle, utilizing split SpCas9 AAV constructs has been attempted[189], even if presenting compromises in activity[190,191]. Achieving dual AAV delivery of SpCas9 and sgRNA is possible[192], but it necessitates the utilization of highly minimized promoters, restricting both expression and tissue specificity. Additionally, the adoption of dual AAV formats involves considerable expenses and presents limitations in co-transduction[193].

Exploring the vast diversity of CRISPR systems present in nature is an efficient way to expand the toolbox. Microbiome databases offer a rich source of genetic information, holding the potential to uncover novel CRISPR-Cas9 systems with unique properties. By mining these databases, researchers can identify systems with distinct PAM specificities, and different target recognition mechanisms, optimizing their performance across a variety of genetic contexts.

Here, a group of 11 Cas9 orthologs with promising properties for genome editing was uncovered starting from the databases presented by Pasolli *et al*[2,152]. Cas9 sizes span from 978 to 1079 aa and all the proteins belong to the subtype II-C, which presents a higher number of <1100 aa Cas9 compared to II-A and II-B. CoCas9 belongs to a poorly characterized species of the *Collinsella* genus, which is especially known for *Collinsella aerofaciens*, the most abundant actinobacterium in the gastrointestinal tract of healthy humans[194–196]. CoCas9 exhibited superior editing efficiency in comparison to the other identified orthologs and SpCas9 when evaluated in mammalian cell lines. CoCas9 is 26.6% smaller than SpCas9 and possesses a combination of 3 PAM motifs (5'-N_4GWNT, 5'-N_4GCDT, and 5'-N_4ATDT): even if complex, this PAM is present in 11.8% of sites in the human genome, which is comparable to 5'-NGG (10.0% of sites). Even presenting a slightly lower editing efficiency compared to SpCas9, CoCas9 showed a remarkable precision profile through GUIDE-seq, highlighting a natural high-fidelity characteristic. The nickase nCoCas9 was used to generate CoABE8e, which reached up to 55% of A>G conversion, and was tested side-by-side with SpABE8e, demonstrating how CoABE8e is, in some cases, able to match SpABE8e with a broader editing window. To prove the advantages offered by CoCas9 compared to the Cas9s collection with similar properties (SaCas9, CjCas9, and Nme2Cas9) we performed a side by side comparative analysis. While SaCas9 is a short representative of II-A, CjCas9, and Nme2Cas9 belong to type II-C and share with CoCas9 30.9% and 31.5% of amino acid sequence similarity, respectively. CoCas9 editing activity is similar to SaCas9 and higher than Nme2Cas9 and CjCas9, but CoCas9 exhibited greater precision than SaCas9, making it preferable for its specificity and targeting range. CoABE8e was tested with other small ABE8es too, which are SaABE8e, CjABE8e, Nme2ABE8e, and SauriABE8e. Data showed that CoABE8e is a valid and competitive base editor and, thanks to PAM preferences, 10.2% of human G>A pathogenic mutations can be targeted only using CoABE8e. CoCas9 consistently produced at least 80% of indels in nearly all edited loci in two relevant primary cell models like HSFs and in HBEs, when delivered as a lentiviral vector. Lastly, CoCas9 and CoABE8e were delivered as single AAVs in HEK293T, showing up to 41% of indels in the first case, and 64.2% of A>G conversion in the second. AAV8-CoCas9 was exploited to edit the mouse retina at the *RHO* gene, generating up to 35% of indels. This outcome is especially significant considering that, despite the retina

being the initial tissue addressed by an approved gene therapy for inherited disorders in the USA[197], it poses certain challenges that researchers are actively working to overcome. To optimize gene editing approaches for a complex organ like the eye, it is crucial to address primary constraints related to long-term treatment impacts. Lower immunogenicity, improved targeting and manufacturing, and enhanced AAV design, such as employing an all-in-one construct, can contribute significantly[198].

In conclusion, the natural diversity of novel Cas9 orthologs identified from metagenomic data has proven to be a valuable resource for expanding the genome editing toolbox. The discovery of CoCas9 and its properties, coupled with the exploration of base editing activities and AAV delivery, demonstrates the potential of these systems for addressing the complexity of gene therapy applications.

As part of future considerations, we aim to investigate whether Cas9 orthologs from commensal species within the human microbiome exhibit reduced immunogenicity compared to Cas9s from pathogenic bacterial strains. Additionally, since this study centers on comparing CoCas9 with wt orthologs, the performance of CoCas9 engineered variants remains a subject for evaluation. Finally, we recently published a computational pipeline to predict PAM sequences of CRISPR-Cas systems from microbial metagenomes, a prediction able to choose the Cas of interest before arriving at the bench. As more metagenomic data becomes available, the existing disparities between *in silico*- and *in vitro*-determined PAMs are expected to diminish, allowing the Cas selection based on the mutation to correct and the PAM requirements[199].

Overall, the expansion of the CRISPR-Cas9 toolbox is crucial for advancing the field of genome editing. By continuously enriching the CRISPR-Cas9 toolbox, researchers can unlock new possibilities for therapeutic interventions, basic research, and biotechnological applications, propelling the field toward greater efficacy and reliability.

Bibliography

1. Pedrazzoli, E. *et al.* An optimized SpCas9 high-fidelity variant for direct protein delivery. *Mol. Ther.* **31**, 2257–2265 (2023).
2. Pedrazzoli, E. *et al.* CoCas9 is a compact nuclease from the human microbiome for efficient and precise genome editing. *Nat. Commun.* **15**, 3478 (2024).
3. Urnov, F. D. Genome Editing B.C. (Before CRISPR): Lasting Lessons from the 'Old Testament'. *CRISPR J* **1**, 34–46 (2018).
4. Khalil, A. M. The genome editing revolution: review. *J. Genet. Eng. Biotechnol.* **18**, 68 (2020).
5. Rouet, P., Smih, F. & Jasin, M. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol. Cell. Biol.* **14**, 8096–8106 (1994).
6. Robb, G. B. Genome editing with CRISPR‐Cas: An overview. *Curr. Protoc. Essent. Lab. Tech.* **19**, (2019).
7. Lieber, M. R. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu. Rev. Biochem.* **79**, 181–211 (2010).
8. Mehta, A. & Haber, J. E. Sources of DNA double-strand breaks and models of recombinational DNA repair. *Cold Spring Harb. Perspect. Biol.* **6**, a016428 (2014).
9. Pannunzio, N. R., Watanabe, G. & Lieber, M. R. Nonhomologous DNA end-joining for repair of DNA double-strand breaks. *J. Biol. Chem.* **293**, 10512–10523 (2018).
10. Scully, R., Panday, A., Elango, R. & Willis, N. A. DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nat. Rev. Mol. Cell Biol.* **20**, 698–714 (2019).
11. Liang, F., Han, M., Romanienko, P. J. & Jasin, M. Homology-directed repair is a major double-strand break repair pathway in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 5172–5177 (1998).
12. Wright, W. D., Shah, S. S. & Heyer, W. D. Homologous recombination and the repair of DNA double-strand breaks. *Journal of Biological Chemistry* vol. 293 10524–10535 Preprint at https://doi.org/10.1074/jbc.TM118.000372 (2018).
13. Azhagiri, M. K. K., Babu, P., Venkatesan, V. & Thangavel, S. Homology-directed gene-editing approaches for hematopoietic stem and progenitor cell gene therapy. *Stem Cell Res. Ther.* **12**, 500 (2021).
14. Silva, G. *et al.* Meganucleases and other tools for targeted genome engineering: perspectives and challenges for gene therapy. *Curr. Gene Ther.* **11**, 11–27 (2011).
15. Tröder, S. E. & Zevnik, B. History of genome editing: From meganucleases to CRISPR. *Lab. Anim.* **56**, 60–68 (2022).
16. Khan, S. H. Genome-Editing Technologies: Concept, Pros, and Cons of Various Genome-Editing Techniques and Bioethical Concerns for Clinical Application. *Mol. Ther. Nucleic Acids* **16**, 326–334 (2019).
17. Belfort, M. & Bonocora, R. P. Homing Endonucleases: From Genetic Anomalies to Programmable Genomic Clippers. in *Homing Endonucleases: Methods and Protocols* (ed. Edgell, D. R.) 1–26 (Humana Press, Totowa, NJ, 2014).

18. Ashworth, J. *et al.* Computational redesign of endonuclease DNA binding and cleavage specificity. *Nature* **441**, 656–659 (2006).
19. Epinat, J.-C. *et al.* A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Res.* **31**, 2952–2962 (2003).
20. Kim, Y. G., Cha, J. & Chandrasegaran, S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 1156–1160 (1996).
21. Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S. & Gregory, P. D. Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.* **11**, 636–646 (2010).
22. Mandell, J. G. & Barbas, C. F., 3rd. Zinc Finger Tools: custom DNA-binding domains for transcription factors and nucleases. *Nucleic Acids Res.* **34**, W516–23 (2006).
23. Matsumoto, D. & Nomura, W. The history of genome editing: advances from the interface of chemistry & biology. *Chem. Commun.* **59**, 7676–7684 (2023).
24. Christian, M. *et al.* Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* **186**, 757–761 (2010).
25. Zhang, F. *et al.* Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat. Biotechnol.* **29**, 149–153 (2011).
26. Li, T. *et al.* TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Res.* **39**, 359–372 (2011).
27. Mahfouz, M. M. *et al.* De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 2623–2628 (2011).
28. Kim, H. & Kim, J.-S. A guide to genome engineering with programmable nucleases. *Nat. Rev. Genet.* **15**, 321–334 (2014).
29. Ain, Q. U., Chung, J. Y. & Kim, Y.-H. Current and future delivery systems for engineered nucleases: ZFN, TALEN and RGEN. *J. Control. Release* **205**, 120–127 (2015).
30. Gaj, T., Gersbach, C. A. & Barbas, C. F. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* **31**, 397–405 (2013).
31. Lu, Y., Happi Mbakam, C., Song, B., Bendavid, E. & Tremblay, J.-P. Improvements of nuclease and nickase gene modification techniques for the treatment of genetic diseases. *Front Genome Ed* **4**, 892769 (2022).
32. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakata, A. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J. Bacteriol.* **169**, 5429–5433 (1987).
33. Mojica, F. J., Juez, G. & Rodríguez-Valera, F. Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI sites. *Mol. Microbiol.* **9**, 613–621 (1993).
34. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* **60**, 174–182 (2005).
35. Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S. D. Clustered regularly interspaced short

palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**, 2551–2561 (2005).

36. Pourcel, C., Salvignol, G. & Vergnaud, G. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* **151**, 653–663 (2005).
37. Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I. & Koonin, E. V. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct* **1**, 7 (2006).
38. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007).
39. Brouns, S. J. J. *et al.* Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960–964 (2008).
40. Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602–607 (2011).
41. Jinek, M. *et al.* A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816–822 (2012).
42. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E2579–86 (2012).
43. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
44. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
45. Jinek, M. *et al.* RNA-programmed genome editing in human cells. *Elife* **2**, e00471 (2013).
46. Thomas, C. M. & Nielsen, K. M. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat. Rev. Microbiol.* **3**, 711–721 (2005).
47. McGinn, J. & Marraffini, L. A. Molecular mechanisms of CRISPR-Cas spacer acquisition. *Nat. Rev. Microbiol.* **17**, 7–12 (2019).
48. Horvath, P. & Barrangou, R. CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327**, 167–170 (2010).
49. Marraffini, L. A. The CRISPR-Cas system of Streptococcus pyogenes: function and applications. in *Streptococcus pyogenes: Basic Biology to Clinical Manifestations* (eds. Ferretti, J. J., Stevens, D. L. & Fischetti, V. A.) (2016).
50. Fineran, P. C. & Charpentier, E. Memory of viral infections by CRISPR-Cas adaptive immune systems: Acquisition of new information. *Virology* **434**, 202–209 (2012).
51. Richter, C., Chang, J. T. & Fineran, P. C. Function and regulation of clustered regularly interspaced short palindromic repeats (CRISPR) / CRISPR associated (Cas) systems. *Viruses* **4**, 2291–2311 (2012).
52. Alkhnbashi, O. S. *et al.* Characterizing leader sequences of CRISPR loci. *Bioinformatics* **32**, i576–i585 (2016).

53. Pougach, K. *et al.* Transcription, processing and function of CRISPR cassettes in Escherichia coli. *Mol. Microbiol.* **77**, 1367–1379 (2010).
54. Jansen, R., van Embden, J. D. A., Gaastra, W. & Schouls, L. M. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* **43**, 1565–1575 (2002).
55. Makarova, K. S. & Koonin, E. V. Annotation and Classification of CRISPR-Cas Systems. in *CRISPR* vol. 1311 47–75 (2015).
56. Makarova, K. S. *et al.* Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* **18**, 67–83 (2020).
57. Pickar-Oliver, A. *et al.* Targeted transcriptional modulation with type I CRISPR-Cas systems in human cells. *Nat. Biotechnol.* **37**, 1493–1501 (2019).
58. He, L., St John James, M., Radovcic, M., Ivancic-Bace, I. & Bolt, E. L. Cas3 Protein-A Review of a Multi-Tasking Machine. *Genes* **11**, (2020).
59. Kolesnik, M. V., Fedorova, I., Karneyeva, K. A., Artamonova, D. N. & Severinov, K. V. Type III CRISPR-Cas Systems: Deciphering the Most Complex Prokaryotic Immune System. *Biochemistry* **86**, 1301–1314 (2021).
60. Moya-Beltrán, A. *et al.* Evolution of Type IV CRISPR-Cas Systems: Insights from CRISPR Loci in Integrative Conjugative Elements of Acidithiobacillia. *CRISPR J* **4**, 656–672 (2021).
61. Hillary, V. E. & Ceasar, S. A. A Review on the Mechanism and Applications of CRISPR/Cas9/Cas12/Cas13/Cas14 Proteins Utilized for Genome Engineering. *Mol. Biotechnol.* **65**, 311–325 (2023).
62. Heler, R. *et al.* Cas9 specifies functional viral targets during CRISPR–Cas adaptation. *Nature* **519**, 199–202 (2015).
63. Wei, Y., Terns, R. M. & Terns, M. P. Cas9 function and host genome sampling in Type II-A CRISPR-Cas adaptation. *Genes Dev.* **29**, 356–361 (2015).
64. Heler, R. *et al.* Mutations in Cas9 Enhance the Rate of Acquisition of Viral Spacer Sequences during the CRISPR-Cas Immune Response. *Mol. Cell* **65**, 168–175 (2017).
65. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733–740 (2009).
66. Mir, A., Edraki, A., Lee, J. & Sontheimer, E. J. Type II-C CRISPR-Cas9 Biology, Mechanism, and Application. *ACS Chem. Biol.* **13**, 357–365 (2018).
67. Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* **13**, 722–736 (2015).
68. Plagens, A., Richter, H., Charpentier, E. & Randau, L. DNA and RNA interference mechanisms by CRISPR-Cas surveillance complexes. *FEMS Microbiol. Rev.* **39**, 442–463 (2015).
69. Kieper, S. N. *et al.* Cas4 Facilitates PAM-Compatible Spacer Selection during CRISPR Adaptation. *Cell Rep.* **22**, 3377–3384 (2018).
70. Leonova, E. I. & Gainetdinov, R. R. CRISPR/Cas9 technology in translational biomedicine. *Cell. Physiol. Biochem.* **54**, 354–370 (2020).
71. Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell*

156, 935–949 (2014).
72. Palermo, G. *et al.* Key role of the REC lobe during CRISPR-Cas9 activation by 'sensing', 'regulating', and 'locking' the catalytic HNH domain. *Q. Rev. Biophys.* **51**, (2018).
73. Killelea, T. & Bolt, E. L. CRISPR-Cas adaptive immunity and the three Rs. *Biosci. Rep.* **37**, 1–9 (2017).
74. Gleditzsch, D. *et al.* PAM identification by CRISPR-Cas effector complexes: diversified mechanisms and structures. *RNA Biol.* **16**, 504–517 (2019).
75. Swarts, D. C., Mosterd, C., van Passel, M. W. J. & Brouns, S. J. J. CRISPR Interference Directs Strand Specific Spacer Acquisition. *PLoS One* **7**, e35888 (2012).
76. Jackson, S. A. *et al.* CRISPR-Cas: Adapting to change. *Science* **356**, eaal5056 (2017).
77. Mohanraju, P. *et al.* Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science* **353**, (2016).
78. Semenova, E. *et al.* Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proceedings of the National Academy of Sciences* **108**, 10098–10103 (2011).
79. Sternberg, S. H., LaFrance, B., Kaplan, M. & Doudna, J. A. Conformational control of DNA target cleavage by CRISPR–Cas9. *Nature* **527**, 110–113 (2015).
80. Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* **31**, (2013).
81. Han, W. & She, Q. CRISPR History: Discovery, Characterization, and Prosperity. in *Progress in Molecular Biology and Translational Science* vol. 152 1–21 (Elsevier Inc., 2017).
82. Hunt, J. M. T., Samson, C. A., Rand, A. du & Sheppard, H. M. Unintended CRISPR-Cas9 editing outcomes: a review of the detection and prevalence of structural variants generated by gene-editing in human cells. *Hum. Genet.* **142**, 705–720 (2023).
83. Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* **36**, 765–771 (2018).
84. Leibowitz, M. L. *et al.* Chromothripsis as an on-target consequence of CRISPR-Cas9 genome editing. *Nat. Genet.* **53**, 895–905 (2021).
85. Turchiano, G. *et al.* Quantitative evaluation of chromosomal rearrangements in gene-edited human stem cells by CAST-Seq. *Cell Stem Cell* **28**, 1136–1147.e5 (2021).
86. Tsuchida, C. A. *et al.* Mitigation of chromosome loss in clinical CRISPR-Cas9-engineered T cells. *Cell* **186**, 4567–4582.e20 (2023).
87. Cheng, A. W. *et al.* Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Res.* **23**, 1163–1171 (2013).
88. Gilbert, L. A. *et al.* Resource CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. *Cell* **154**, 442–451 (2012).
89. Amabile, A. *et al.* Inheritable Silencing of Endogenous Genes by Hit-and-Run Targeted Epigenetic Editing. *Cell* **167**, 219–232.e14 (2016).
90. Chen, M. & Qi, L. S. Repurposing CRISPR System for Transcriptional Activation. in *RNA Activation* (ed. Li, L.-C.) 147–157 (Springer Singapore, Singapore, 2017).

91. Anzalone, A. V. *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, (2019).
92. Chen, P. J. *et al.* Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. *Cell* **184**, 5635–5652.e29 (2021).
93. Doman, J. L. *et al.* Phage-assisted evolution and protein engineering yield compact, efficient prime editors. *Cell* **186**, 3983–4002.e26 (2023).
94. Huang, Z. & Liu, G. Current advancement in the application of prime editing. *Front Bioeng Biotechnol* **11**, 1039315 (2023).
95. Gu, S., Bodai, Z., Cowan, Q. T. & Komor, A. C. Base editors: Expanding the types of DNA damage products harnessed for genome editing. *Gene and Genome Editing* **1**, 100005 (2021).
96. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
97. Gaudelli, N. M. *et al.* Programmable base editing of A-T to G-C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).
98. Jiang, F. & Doudna, J. A. CRISPR-Cas9 Structures and Mechanisms. *Annu. Rev. Biophys.* **46**, 505–529 (2017).
99. Anzalone, A. V., Koblan, L. W. & Liu, D. R. Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nature Biotechnology* vol. 38 824–844 Preprint at https://doi.org/10.1038/s41587-020-0561-9 (2020).
100. Komor, A. C. *et al.* Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. *Science Advances* **3**, 1–9 (2017).
101. Koblan, L. W. *et al.* Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat. Biotechnol.* **36**, 843–846 (2018).
102. Rees, H. A. & Liu, D. R. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.* **19**, 770–788 (2018).
103. Wolf, J., Gerber, A. P. & Keller, W. tadA, an essential tRNA-specific adenosine deaminase from Escherichia coli. *EMBO J.* **21**, 3841–3851 (2002).
104. Molla, K. A. & Yang, Y. CRISPR/Cas-Mediated Base Editing: Technical Considerations and Practical Applications. *Trends Biotechnol.* **37**, 1121–1142 (2019).
105. Hu, J. H. *et al.* Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57–63 (2018).
106. Yang, L. *et al.* Increasing targeting scope of adenosine base editors in mouse and rat embryos through fusion of TadA deaminase with Cas9 variants. *Protein Cell* **9**, 814–819 (2018).
107. Nishimasu, H. *et al.* Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science* **361**, 1259–1262 (2018).
108. Richter, M. F. *et al.* Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat. Biotechnol.* **38**, 883–891 (2020).
109. Grünewald, J. *et al.* CRISPR DNA base editors with reduced RNA off-target and self-editing activities. *Nat. Biotechnol.* **37**, 1041–1048 (2019).

110. Zhao, D. *et al.* Glycosylase base editors enable C-to-A and C-to-G base changes. *Nat. Biotechnol.* **39**, 35–40 (2021).
111. Tong, H. *et al.* Programmable A-to-Y base editing by fusing an adenine base editor with an N-methylpurine DNA glycosylase. *Nat. Biotechnol.* **41**, 1080–1084 (2023).
112. Chen, L. *et al.* Adenine transversion editors enable precise, efficient A•T-to-C•G base editing in mammalian cells and embryos. *Nat. Biotechnol.* (2023) doi:10.1038/s41587-023-01821-9.
113. Walton, R. T., Christie, K. A., Whittaker, M. N. & Kleinstiver, B. P. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* **368**, 290–296 (2020).
114. Tu, T. *et al.* A precise and efficient adenine base editor. *Mol. Ther.* **30**, 2933–2941 (2022).
115. Porto, E. M. & Komor, A. C. In the business of base editors: Evolution from bench to bedside. *PLoS Biol.* **21**, e3002071 (2023).
116. Wang, J. Y. & Doudna, J. A. CRISPR technology: A decade of genome editing is only the beginning. *Science* **379**, eadd8643 (2023).
117. Yeh, W.-H., Chiang, H., Rees, H. A., Edge, A. S. B. & Liu, D. R. In vivo base editing of post-mitotic sensory cells. *Nat. Commun.* **9**, 2184 (2018).
118. Zhang, X.-H., Tee, L. Y., Wang, X.-G., Huang, Q.-S. & Yang, S.-H. Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. *Molecular Therapy - Nucleic Acids* **4**, e264 (2015).
119. Kleinstiver, B. P. *et al.* High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490–495 (2016).
120. Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84–88 (2016).
121. Chen, J. S. *et al.* Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* **550**, 407–410 (2017).
122. Bratovič, M. *et al.* Bridge helix arginines play a critical role in Cas9 sensitivity to mismatches. *Nat. Chem. Biol.* (2020) doi:10.1038/s41589-020-0490-4.
123. Bravo, J. P. K. *et al.* Structural basis for mismatch surveillance by CRISPR–Cas9. *Nature* 1–5 (2022).
124. Casini, A. *et al.* A highly specific SpCas9 variant is identified by in vivo screening in yeast. *Nat. Biotechnol.* **36**, 265–271 (2018).
125. Lee, J. K. *et al.* Directed evolution of CRISPR-Cas9 to increase its specificity. *Nat. Commun.* **9**, (2018).
126. Tao, J., Bauer, D. E. & Chiarle, R. Assessing and advancing the safety of CRISPR-Cas tools: from DNA to RNA editing. *Nature Communications 2023 14:1* **14**, 1–16 (2023).
127. Lundstrom, K. Viral Vectors in Gene Therapy. *Diseases* **6**, 42 (2018).
128. Ricobaraza, A., Gonzalez-Aparicio, M., Mora-Jimenez, L., Lumbreras, S. & Hernandez-Alcoceba, R. High-Capacity Adenoviral Vectors: Expanding the Scope of Gene Therapy. *Int. J. Mol. Sci.* **21**, (2020).
129. Poletti, V. & Mavilio, F. Designing Lentiviral Vectors for Gene Therapy of Genetic Diseases. *Viruses* **13**, (2021).

130. Soldi, M. *et al.* Laboratory-Scale Lentiviral Vector Production and Purification for Enhanced Ex Vivo and In Vivo Genetic Engineering. *Mol Ther Methods Clin Dev* **19**, 411–425 (2020).
131. Wanisch, K. & Yáñez-Muñoz, R. J. Integration-deficient lentiviral vectors: a slow coming of age. *Mol. Ther.* **17**, 1316–1332 (2009).
132. Colella, P., Ronzitti, G. & Mingozzi, F. Emerging Issues in AAV-Mediated In Vivo Gene Therapy. *Mol Ther Methods Clin Dev* **8**, 87–104 (2018).
133. Liang, X. *et al.* Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. *J. Biotechnol.* **208**, 44–53 (2015).
134. Mitchell, M. J. *et al.* Engineering precision nanoparticles for drug delivery. *Nat. Rev. Drug Discov.* **20**, 101–124 (2021).
135. Kazemian, P. *et al.* Lipid-Nanoparticle-Based Delivery of CRISPR/Cas9 Genome-Editing Components. *Mol. Pharm.* **19**, 1669–1686 (2022).
136. Kleinstiver, B. P. *et al.* Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481–485 (2015).
137. Ran, F. A. *et al.* In vivo genome editing using Staphylococcus aureus Cas9. *Nature* **520**, 186–191 (2015).
138. Cebrian-Serrano, A. & Davies, B. CRISPR-Cas orthologues and variants: optimizing the repertoire, specificity and delivery of genome engineering tools. *Mamm. Genome* **28**, 247–261 (2017).
139. Kim, E. *et al.* In vivo genome editing with a small Cas9 orthologue derived from Campylobacter jejuni. *Nat. Commun.* **8**, 1–12 (2017).
140. Koo, T. *et al.* Functional Rescue of Dystrophin Deficiency in Mice Caused by Frameshift Mutations Using Campylobacter jejuni Cas9. *Mol. Ther.* **26**, 1529–1538 (2018).
141. Jo, D. H. *et al.* Long-Term Effects of In Vivo Genome Editing in the Mouse Retina Using Campylobacter jejuni Cas9 Expressed via Adeno-Associated Virus. *Mol. Ther.* **27**, 130–136 (2019).
142. Zhang, Y. *et al.* Processing-Independent CRISPR RNAs Limit Natural Transformation in Neisseria meningitidis. *Mol. Cell* **50**, 488–503 (2013).
143. Esvelt, K. M. *et al.* Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat. Methods* **10**, 1116–1121 (2013).
144. Edraki, A. *et al.* A Compact, High-Accuracy Cas9 with a Dinucleotide PAM for In Vivo Genome Editing. *Mol. Cell* **73**, 714–726.e4 (2019).
145. Lee, C. M., Cradick, T. J. & Bao, G. The neisseria meningitidis CRISPR-Cas9 system enables specific genome editing in mammalian cells. *Mol. Ther.* **24**, 645–654 (2016).
146. Hu, Z. *et al.* A compact Cas9 ortholog from Staphylococcus Auricularis (SauriCas9) expands the DNA targeting scope. *PLoS Biol.* **18**, e3000686 (2020).
147. Vakulskas, C. A. *et al.* A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells. *Nat. Med.* **24**, 1216–1224 (2018).
148. Schmid-Burgk, J. L. *et al.* Highly Parallel Profiling of Cas9 Variant Specificity. *Mol. Cell* **78**,

794–800.e8 (2020).

149. Kim, N. *et al.* Prediction of the sequence-specific cleavage activity of Cas9 variants. *Nat. Biotechnol.* **38**, 1328–1336 (2020).
150. Kim, S., Kim, D., Cho, S. W., Kim, J. & Kim, J.-S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res.* **24**, 1012–1019 (2014).
151. Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**, 187–197 (2014).
152. Pasolli, E. *et al.* Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. *Cell* **176**, 649–662.e20 (2019).
153. Bland, C. *et al.* CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* **8**, 209 (2007).
154. Russel, J., Pinilla-Redondo, R., Mayo-Muñoz, D., Shah, S. A. & Sørensen, S. J. CRISPRCasTyper: Automated Identification, Annotation, and Classification of CRISPR-Cas Loci. *CRISPR J* **3**, 462–469 (2020).
155. Chyou, T.-Y. & Brown, C. M. Prediction and diversity of tracrRNAs from type II CRISPR-Cas systems. *RNA Biol.* **16**, 423–434 (2019).
156. Gardner, P. P., Barquist, L., Bateman, A., Nawrocki, E. P. & Weinberg, Z. RNIE: genome-wide prediction of bacterial intrinsic terminators. *Nucleic Acids Res.* **39**, 5845–5852 (2011).
157. Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **6**, 26 (2011).
158. Briner, A. E. *et al.* Guide RNA functional modules direct Cas9 activity and orthogonality. *Mol. Cell* **56**, 333–339 (2014).
159. Karvelis, T., Young, J. K. & Siksnys, V. A pipeline for characterization of novel Cas9 orthologs. *Methods Enzymol.* **616**, 219–240 (2019).
160. Tareen, A. & Kinney, J. B. Logomaker: beautiful sequence logos in Python. *Bioinformatics* **36**, 2272–2274 (2020).
161. Walton, R. T., Hsu, J. Y., Joung, J. K. & Kleinstiver, B. P. Scalable characterization of the PAM requirements of CRISPR-Cas enzymes using HT-PAMDA. *Nat. Protoc.* **16**, 1511–1547 (2021).
162. Landrum, M. J. *et al.* ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* **46**, D1062–D1067 (2018).
163. Tornabene, P. *et al.* Inclusion of a degron reduces levels of undesired inteins after AAV-mediated protein trans-splicing in the retina. *Molecular Therapy - Methods & Clinical Development* **23**, 448–459 (2021).
164. Scudieri, P. *et al.* Association of TMEM16A chloride channel overexpression with airway goblet cell metaplasia. *J. Physiol.* **590**, 6141–6155 (2012).
165. Scudieri, P. *et al.* Ionocytes and CFTR Chloride Channel Expression in Normal and Cystic Fibrosis Nasal and Bronchial Epithelial Cells. *Cells* **9**, (2020).
166. Mou, H. *et al.* Dual SMAD Signaling Inhibition Enables Long-Term Expansion of Diverse Epithelial Basal Cells. *Cell Stem Cell* **19**, 217–231 (2016).

167. Pizzato, M. *et al.* A one-step SYBR Green I-based product-enhanced reverse transcriptase assay for the quantitation of retroviruses in cell culture supernatants. *J. Virol. Methods* **156**, 1–7 (2009).
168. Kimura, T. *et al.* Production of adeno-associated virus vectors for in vitro and in vivo applications. *Sci. Rep.* **9**, 1–13 (2019).
169. Aurnhammer, C. *et al.* Universal real-time PCR for the detection and quantification of adeno-associated virus serotype 2-derived inverted terminal repeat sequences. *Hum. Gene Ther. Methods* **23**, 18–28 (2012).
170. Doria, M., Ferrara, A. & Auricchio, A. AAV2/8 Vectors Purified from Culture Medium with a Simple and Rapid Protocol Transduce Murine Liver, Muscle, and Retina Efficiently. *Hum. Gene Ther. Methods* **24**, 392–398 (2013).
171. Robichaux, M. A. *et al.* Subcellular localization of mutant P23H rhodopsin in an RFP fusion knock-in mouse model of retinitis pigmentosa. *Dis. Model. Mech.* **15**, (2022).
172. Liang, F. Q., Anand, V., Maguire, A. M. & Bennett, J. Intraocular delivery of recombinant virus. *Methods Mol. Med.* **47**, 125–139 (2001).
173. Nobles, C. L. *et al.* iGUIDE: an improved pipeline for analyzing CRISPR cleavage specificity. *Genome Biol.* **20**, 14 (2019).
174. Davis, J. R. *et al.* Efficient in vivo base editing via single adeno-associated viruses with size-optimized genomes encoding compact adenine base editors. *Nature Biomedical Engineering* **6**, 1272–1283 (2022).
175. Li, P. *et al.* Allele-Specific CRISPR-Cas9 Genome Editing of the Single-Base P23H Mutation for Rhodopsin-Associated Dominant Retinitis Pigmentosa. *CRISPR J* **1**, 55–64 (2018).
176. Patrizi, C. *et al.* Allele-specific editing ameliorates dominant retinitis pigmentosa in a transgenic mouse model. *Am. J. Hum. Genet.* **108**, 295–308 (2021).
177. Wu, W.-H. *et al.* CRISPR genome surgery in a novel humanized model for autosomal dominant retinitis pigmentosa. *Mol. Ther.* **30**, 1407–1420 (2022).
178. Wu, X. *et al.* Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat. Biotechnol.* **32**, 670–676 (2014).
179. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80–84 (2014).
180. Okada, K. *et al.* Key sequence features of CRISPR RNA for dual-guide CRISPR-Cas9 ribonucleoprotein complexes assembled with wild-type or HiFi Cas9. *Nucleic Acids Res.* **50**, 2854–2871 (2022).
181. Guo, C., Ma, X., Gao, F. & Guo, Y. Off-target effects in CRISPR/Cas9 gene editing. *Front Bioeng Biotechnol* **11**, 1143157 (2023).
182. Guo, M. *et al.* Structural insights into a high fidelity variant of SpCas9. *Cell Research 2019 29:3* **29**, 183–192 (2019).
183. Frati, G. & Miccio, A. Genome Editing for β-Hemoglobinopathies: Advances and Challenges. *J. Clin. Med. Res.* **10**, (2021).
184. Frangoul, H. *et al.* CRISPR-Cas9 Gene Editing for Sickle Cell Disease and β-Thalassemia. *N. Engl. J. Med.* **384**, 252–260 (2021).

185. Huang, X., Yang, D., Zhang, J., Xu, J. & Chen, Y. E. Recent Advances in Improving Gene-Editing Specificity through CRISPR-Cas9 Nuclease Engineering. *Cells* **11**, (2022).
186. Idoko-Akoh, A., Taylor, L., Sang, H. M. & Mcgrew, M. J. High fidelity CRISPR/Cas9 increases precise monoallelic and biallelic editing events in primordial germ cells. doi:10.1038/s41598-018-33244-x.
187. Schmidt-Dannert, C. & Arnold, F. H. Directed evolution of industrial enzymes. *Trends Biotechnol.* **17**, 135–136 (1999).
188. Collias, D. & Beisel, C. L. CRISPR technologies and the search for the PAM-free nuclease. *Nat. Commun.* **12**, 555 (2021).
189. Chew, W. L. *et al.* A multifunctional AAV-CRISPR-Cas9 and its host response. *Nat. Methods* **13**, 868–874 (2016).
190. Wright, A. V. *et al.* Rational design of a split-Cas9 enzyme complex. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 2984–2989 (2015).
191. Zetsche, B., Volz, S. E. & Zhang, F. A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nat. Biotechnol.* **33**, 139–142 (2015).
192. Suzuki, K. *et al.* In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature* **540**, 144–149 (2016).
193. Komor, A. C., Badran, A. H. & Liu, D. R. CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. *Cell* **168**, 20–36 (2017).
194. Kageyama, A., Benno, Y. & Nakase, T. Phylogenetic and phenotypic evidence for the transfer of Eubacterium aerofaciens to the genus Collinsella as Collinsella aerofaciens gen. nov., comb. nov. *Int. J. Syst. Bacteriol.* **49 Pt 2**, 557–565 (1999).
195. Bag, S., Ghosh, T. S. & Das, B. Complete Genome Sequence of Collinsella aerofaciens Isolated from the Gut of a Healthy Indian Subject. *Genome Announc.* **5**, (2017).
196. Bilen, M. *et al.* Genomic and phenotypic description of the newly isolated human species Collinsella bouchesdurhonensis sp. nov. *Microbiologyopen* **7**, e00580 (2018).
197. Trapani, I. & Auricchio, A. Seeing the Light after 25 Years of Retinal Gene Therapy. *Trends Mol. Med.* **24**, 669–681 (2018).
198. Drag, S., Dotiwala, F. & Upadhyay, A. K. Gene Therapy for Retinal Degenerative Diseases: Progress, Challenges, and Future Directions. *Invest. Ophthalmol. Vis. Sci.* **64**, 39 (2023).
199. Ciciani, M. *et al.* Automated identification of sequence-tailored Cas9 proteins using massive metagenomic data. *Nat. Commun.* **13**, 6474 (2022).

Appendix

Figure S1

Figure S1. CoCas9 and SpCas9 comparative off-target analysis*. Off-target sites detected by GUIDE-seq with CoCas9 and SpCas9 targeting the genomic loci HPRT (**a**),*

*VEGFAsite2 (**b**), Chr6 (**c**), ZSCAN2 (**d**). For SpCas9 VEGFAsite2 in (**b**) 135 off-target sites are reported out of a total 1815 identified sites. Each detected off-target is accompanied by its number of GUIDE-seq reads, which is an indirect measure of the propensity of the site to be cleaved by the nucleases. Black squares indicate the ON-target site.*

Table 1. Sequences of all the Cas9 orthologs shown in this work.

CoCas9	MEITINREIGKLGLPRHLVLGMDPGIASCGFALIDTANREILDLGVRLFDSPTHPKTGQSLAVIRRGFRSTRRNIDR TQARLKHCLQILKAYGLIPQDATKEYFHTTKGDKQPLKLRVDGLDRLLNDREWALVLYSLCKRRGYIPHGEGNQD KSSEGGKVLSALAANKEAIAETSCRTVGEWLAQQPQSRNRGGNYDKCVTHAQLIEETHILFDAQRSFGSKYASP EFEAAYIEVCDWERSRKDFDRRTYDLVGHCSYFPTEKRAARCTLTSELVSAYGALGNITIIHDDGTSRALSATERD ECIAILFSCEPIRGNKDCAVKFGALRKALDLSSGDYFKGVPAADEKTREVYKPKGWRVLRNTLNAANPILLQRLR DDRNLADAVMEAVAYSSALPVLQEQLQGLPLSEAEIEALCRLPYSSKALNGYGNRSKKALDMLLDCLEEPEVLNL TQAENDCGLLGLRIAGTQLERSDRLMPYETWIERTGRTNNNPVVIRAMSQMRKVVNAICRKWGVPNEIHVELDR ELRLPQRAKDEIAKANKKNEKNRERIAGQIAELRGCTADEVTGKQIEKYRLWEEQECFDLYTGAKIEVDRLISDDT YTQIDHILPFSRTGENSRNNKVLVLAKSNQDKREQTPYEWMSHDGAPSWDAFERRVQENQKLSRRKKNFLLEK DLDTKEGEFLARSFTDTAYMSREVCAYLADCLLFPDDGAKAHVVPTTGRATAWLRRRWGLNFGSNGEKDRSDD RHHATDACVIAACSRSLVIKTARINQETHWSITRGMNETQRRDAIMKALESVMPWETFANEVRAAHDFVVPTRFV PRKGKGELFEQTVYRYAGVNAQGKDIARKASSDKDIVMGNAVVSADEKSVIKVSEMLCLRLWHDPEAKKGQGA WYADPVYKADIPALKDGTYVPRIAKQKYGRKVWKAVPNSALTQKPLEIYLGDLIKVGDKLGRYNGYNIATANWSF VDALTKKEIAFPSVGMLSNELQPIIIRESILDN
BsCas9	MKHPYGIGLDIGIASVGWAVVALNENAEPYGLIRCGSRIFDKAEQPKTGDSLAAPRREARSARRRLRRRSLRKA DLYELMEKNGLPGKAEIEQAVQAGHLPDVYALRVQALDGPVTALDFARILLHLMQRRGFRSNRKADDAQKDGKL LQAIDANTRRMEANRYRTVGEMMYRDPVFAEHKRNKAENYLSTVKRDQIIDEVRLVFAAQRQYGATWASPEME AEYLCILTRQRSFAEGPGKGSPYSGSNRVGTCTLEGKSEQRAAKAAFSFEYFTLLQKINHIRIAENGTSRTLTPAE RQILLSVCCQTDKMDFARVRKALALPEEARFNMVRYRGEQTAEACEKKEKITALPCYHKMRKALNTLRKDHIRNI SREQLDAAGAALTNPENEDKLREALKQAEFEPLEIEALLTLPSFAGYGHISVKACRKLIPYLEQGMNYNDACQAA GYDFQGRQNGEKAQFLPASTEEMDDITSPVVRRAVAQTIKVVNAIIREQGESPVSIHLELAREMNKNFQQRSELD KAMRDNSAENERLMKELNELFPGRTVTGQDLVKYRLWKEQNEICAYSLEKLDLTRVITEAGYAEVDHIVPYSISF DDRRTNKVLVLASENRQKGNRLPLQYLQGKRRDDFIVYTKANVKNFRKRQNLLKERLSEEDGKGYIQRNLQDT QYIAAFMLNYIRNHLAFADCSGAGKRRVVAVNGAVTAFLRKRWGLSKVRTDGDLHHAADAAVIACTTQGMIKRV SDFCKRAETTAVRNEHFPEPWPRFRDELTQRLSACPQEDLMKINPVYYATVDISSIQPVFVSRMPRHKVTGAAH EETIRSQVNEKYTAVRKSITELSLDKDGKIKDYFKPSSDTLLYEALKKRLTEFGGNAKKAFAEPFYKPRADGTPGA QVRKVKVVAKMNNTIPVRSGGGVAKGGDMIRIDVYYVPGDGYYWVPIYVADTVKETLPNKAVVRDKTMEEWKE MKEDDFLFSLYSNDLILVERDKPICFSLMHEDSTLPKKYETKKELVYYKGGDISNGGIRIETHEGAYFLKSLTFGIV QKVQKYQVDVLGNYTPVKKEKRQTFPAQRR
Al1Cas9	MKKKIFGFDLGIASIGWAVIDHSDENFDPETGEIIEGKVVGCGVRCFPVAENPKDGSSLAAPRREKRLLRRITRRK ARRMLGIKRLFVAKGLAASTAELETLYAAQTGGDVWNLRAEALRRPLSKEELLRVLTHLAKHRGFKSYRKAAEEA DKESGRILTAIAENRKETAGFQTLAQMIVERAKHSDDHKMRNYTSQEGENKGVAVYVNSIPREEIEKETKLIFEYQ KQFGLFTEDLYRDFCKIASRYREAGSVGHMVGRCRFEPEQPRAPKEAPSAELFVALSKINNLKVTVDGERRFLN GEERKALLELLKNTKEVKYLTIKNKLFKGREVFFDDVNYAQKTKKGKSGEEKAVNPEDAKFYAMKGWHKLKAAF SPEQWKEVGSNLPLLDLGMTAVVCEKNDAGIERFLSEKGIPEDYREVFKKLTGSEFINLSLKALYKLNPYLAEGLK YNQACEKAGYDFREDGIKLAEEKGLLLPPIADDKLTTVPVVNRAVAQFRKVYNAMVRTYGAPDQINLEIGRDLKK SRDERNQIMRRQKENEAERKEAEDWLEKEGLAANGKNMLKYRLYRQQNGKCIYSGKAIDLRRLDENGYCDVD HIIPYSRSLDDGQNNKVLCLAEENRKKGSQTPYEYLEPLGRWEEFETVVNTTPSINRYKRNNLLNKDYKEKEND LEFRERNANDNSYIARYVKRYLEDAIDFSASSCTIGNRVQVRTGSLTDYLRHQWGLIKDRDASDRHHAQDAVVV ACATQGMVQKLSKLSAIFENKDDFRRKKAEELGHEEAEAWYKYVKQQIREPWSGFRAEVLASLEKVFVSRPPR KNATGEIHQETIRTVNPKRKKYNEKEILSGIKIRGGLAKNGLMLRTDVFVKKNKKGKDEFYLVPVYLSDMGKELPN KAMVPGKKENEWIELDETCQFKFSFYMDDLIKIKKKENEIFGYFRGTNRATASVSVTTHDRSHTFEGIGVKTQDGI EKYQVDPLGRIAKVKKEIRLPLTMMKKNRHKKEE

Al2Cas9	MRKRIFGFDLGIASIGWAVVDFDKEYFDHETGEVIEGQIIKSGVRCFPVAENPKDGSSLAAPRREKRLARRICRRK ARRMQEIKQLFIDSGLVADSDMLQKLYAEQKDGDVWNLRIKALTQKLTAAELVRVLTHLAKHRGFKSYRKAAEES DTEGGKVLKAIKANSIQLSDNKTLAQIIVERAGKNGKKRNYTETNAKGKEEAVYINSIPRIEIMRETKLIYEAQKGE KGLFTDELYAKFCNIAFRYREAGSVANMVGQCIFEKDEKRAPREAPTSELFVALTKINNMSLNDNGKIRTVSPTER KQILEILKNTKLVKYSTLSSKVFAKGVKFRDIDYNKTEKRNKNNEVKTVNPEDITFYEMEGWHKLKAQFDKDEWA VVSKDTHLLDKVVNIIACEKNDASITKALKELGVKDEWIEKFNNCTFDKFINLSFKALYNIVPYMMDGLHYNEACE KVGYDHKLTVDKLVEHKGIYLEAIASDKLTKVPVVNRTVAQFRKVYNAMARQYGLPDQINIEMGRDLKKTFEERK DLKSLQDENMKQREIIEKELNEHKIKVNAKNILKYRLYKEQSCKCIYSGKTIDLERLDEVGYLDVDHIIPYSRSFDD SYNNKVLCLSEENRKKGNKTPFDYIKNEMAWAEFEARVKLLHNKKKEDLLLCKDFQDRELAFKERNANDNSYIS RYVKQYCEDGIDFSSSPWKDIKNRIQMRAGYLTDYLRWQWGLSKDRNANDRHHAQDAIVIACATQGMVSYLSY VSSVFENKFAVQAKTGEAWYQSLKKKWNEPWTGFRDSVLKSIDEIFVSRPPRKNATGEIHQERIRTINPNNKKYS EKDVKSGIFIRGGLANNGNMLRTDVFVKKNKKGKEQFYLVPVYLSDMGKELPNKAIVAKKDEKDWVIIDDSYTFK FSMFMDDLIKITKGNKAILGYFCGTHRGNASITLEEHDRAKITEGIGVKGLDKFQKFSVDPLGNITEIKQETRLPLT NIKSNKQRMAERKARREKLEQEKQ
SuCas9	MSSKMRYRLALDLGTTSIGWAMLRLDQQDRPCAVIRAGVRIFNDGRVPKTGEALAVQRRLARQQRRTRDRSLR RKNKLINMLVRFGFFPENLAERKSLERLNPYELRAKGLEHQLKPEEFARALFHLAQRRGFKSNRKTDRGDAESG VMKTAIVETRQEIETLGCRTIGEYLYKRLQAGLGTRARLRDKPKVKGGATKSYDLYLERSQVETEFDALWDVQSR FAPELFTQEKREALRDTIFHQRDLRPVKPGRCTLFPEEERAPLALPSQQRFRICQEVNNLRKIDEHLCSIELTQTE KAKLVQALSGARKKTFEQLRKVIGYEGQFNLEEAGAGRADLKGNETAVVIGKIIGKDWLKYSLSEQDKLVLMLLNI ESESELIGTLQEEFDLSEAVAKKLSGAPLVNGYGSLSRKAIDLILPFLEQGMTYDKAVAAADIPSHSDLTWQVKSD ELMDQLPYYGEVLQRHVGFGTGNPEDAPEVRFGKIANPTVHICLNQLRLIVNLIIKRYGNPDGVVIEVARDLKLSR EKKNEIVRKQKKNQEKREQVRKEIAAILECDETLVTSEDIEKWLLWEELNPENALDRRCPYTGRQISAVDLLSKAF EIEHILPRSRTLDDSRANKTVSYWEANRLKSNLTPWEAREIFEAHGWNYEDILKRAENLPPNKRYRFGADGYKT WLREDSDFLARELTDTQYISRIAREYVSLVCPKNTYVIPGRMTAMLRHCLGLNTILSDDDKKNRNDHRHHAVDAC VIGITDRSLLNKIAALSARGEEQNDRKFQRSVEEPWRGFRESVCRAVERINVSHRPEHSYEGQMHDQTAYPIGK NGQGWKKEVKSNGQTEWKKISVIPIVSQKAAQRHGYLSDGTLRPYKGYSGNSNFCIEITVNEKGKWQAEFVTT YDAYQIARKFGKDKLYGKESQSGHKLVMRLMKKDVVAMEDVDGCRRLYLLHKFSTNGVLSFAPINESNVSARVT EGSFKYVSKTAGSLQRAKARQVMLSPLGDRS
ClCas9	MKYGIGLDCGIASVGYCITELDSNDEPKRIVRLGSRIFDKAENPKDGSSLAKPRREARGLRRRIRRHQHRLQRIR YILVSDGIVTQDELDTMFQGQLSDIYEIRARALDEPISNVEFARVLINLAQRRGFKSNRKVDENTKDKETGKLLGAI EKNRENIAKNGYRTVGEMLYKDERYSKYKRNKGEDYLNTVSRDMIADEIKMIFSAQRSFRMPFATEEIEARYTDI VMSQRPFDLGPGEGNEKSPSPYAGNQIEKMVGRCTFFPDEYRAVKASYSFQLFSLLQGINNITLVDDGGNAFPL SQQERNELKEYCFKTKSVTYASIRKKLKISPDFKFKNITYTEKGVEESEKKTKFQHLNIYHQMKSTLGDAMNSLS HDELDEIGRIFTFYKNDTKIIEALEKTDIDKDLYPALLNLPSFSKTGHISVKACKLLIPFLEEGKTYNEACECAGLDFK AHSNAEKKILLPPKSDELDDIVNPVVRRAVSQTVKVVNAIIREMGTSPTYLNIELAQELSKSKKERDEIEKNYLLNR AKNEKIKKEIVDNFGFEPKGQDIVKLKLYHEQDGICPYSLEPIKYDRLFEVGYVDIDHIIPYSACFDDSYNNKVLTFS SENRQKRNRVPLEYLAENKKSDYRVWVNSNIRNYRKKQNLLKEHFTEKDKEDFKQRNLNDTKYLSRVLYNYIND NLIFEDFANGRKRHVISVNGAATAHMRKCWGIDKIRENGDLHHAVDAAVISCVTQGMVNKISKYSFDRETMFDVD YRTGEVKERFPLPYPNFRKELEARSEIEDEKRLKSVLLSFPNYSYDDAESAKPAFVSRMPRRKVTGPAHKDTIRS GKVDGCKISKVELTSLKLKDGAIENYYNPESDTLLYNALLQRLIEFDGDAKQAFAEELHKPKADGTPGPIVKKVKII ERSSSSVRARNENKLGIADNGSMVRIDVFYVENDGYYFVPIYVADTIKPTLPNLACVPGGKPWKEMDDKDFVFS LYQNDLIRITAKKDMKFSVVNKDSTLPANKYENELLLYYGGADISTASINGITHDNSYKFKSCGIKSLVNIEKFTVDP LGNVNKVNKEKRMYFN
Al4Cas9	MSLIFGFDLGIASIGWAAVKMSDPTASDTDAAAPDNRPVGEILGAGVRCFDQAQNAVDRRVARGARRRIRHRAQ RMRDIRKILRQYNLIDIPEPKFNGQNNFYLNRDTDINIWQLRAVDAFTRKLTPRETGRILYHLAKHRGYDDITYPVL TGVVPEKADDAETKEELKAIGAIKDNFAKLHNENPDQTMCQMLYATEHGQMRNGRKSVIKITKDGTQRTEEQSS YSNSIPRSEIKREAEMILHAQCGFGNDFSDVFKEWCVVAFHQLKFNESKSPLAKSIEKMRGKCPFTNKPVAPKES PTAQMFVALTKCRQNGFTVAETNAIIDALYKKKTGLKYADVRKLLKWDDDSQFKTLNYSRTYDKENKKWIEPDVK KIENSKFYAFTGWHKLAKYTTDVATMDKIFDIIVTQKTPENIERAVSGILPEHAAAIAQITSSEFIKLSITALQKIIPEMA AGKKYNEAVADGLGKDFRVQNVSYIDRDDGVSAGCLHQINWNALGNRITSPVAKRTLGQLRKVYNAMVYRFGA PDRIHLEIGRELKNSPDDIARMNREQERNRQATEEARKEYGKNYFKYKLYMEQGGKCPYCGATLVAGDWAAYEI DHILPYSRSLDNSQSNKVLVCHECNQRKGNKTPYEFLTTEQFHDMAVRARALHNPAKFKKLTNTELPQSSDEQN GFIERNANDNATIARFATEYLEYGIDWPESDNIKKRVLVRTGSLTDYLRHQWGLVKNRDESDKHHAQDAIVIACAT QGMVKYLSTLSARFENKYALIEKYGQAWYNVLKTSVQEPWAGFRTAVFSLLDNIIVSRPPRCNATGAAHKETIYA NPKSYRSKRGLNKNRDKGSMKIRHGRVERGDMFRFDIWRDARGKYVCVPIFVADTKGKDDNNFKVPDGEFICS LHKDDYVRIYTSNQVIEGYISQMQEVIHGDKIYRYLALYTHDNHKTIVCKDDKSKAVPEIFRANVSDATNIEKFAMS LLGTPKPIKLPEQRTPVVK

GeCas9	MEKRIFGFDIGIASLGWAVVDFDDTAEPENNIYPTGKIVKSGVRCFPVAENPKDGSSLAFPRRQKRSQRKLCRRK ARRMAGIKSLFVANGLIDRASLFNEKENIYKARDKADVWDLRVKALTEKLTTVEFIRVLTHLAKHRGFKSYRIASEK ADKENGKVLEAVKANRVLLENGKTLAQIIVEKGGKKRNRDDREGKSTYENSIPRDEIERETKLIFEKQRELGLTAA TEKLQRDFVEIAFRFRKINGKKIEKMIGKCLFEKDEPRSPKNAPSAEFFVAWTKINNCHVRDNDGKLRFLTQEEK DAAFELLKNKKEVKYADIRKKLFPNRPDIRFADVEYNPKPVFDKKTGEIKELPEPESLKFFALKGWHDLKSAIDVS AYPNERLDKAVTVIATQKNDADISKELKNLGFSASDVENLAGLSFSKFISLSLTALYKILPEMQAGKKYNEACEAVG YDFKSTGGSFAAQKGKFLPPIPESLATTVPVVNRAMAQFRKVYNALVREYGTPDQINVELARDVYNNHDERREI ADRQNEYAEAKKKVKASACSKLELADISGRDLLKFLLYEQQDGKCVYSGEKLDLHRLVEQDYCDVDHIIPYSRSL DNSQNNKVLCLSRENRQKSDKTPLEYITDPVKQAEFIARVKAMKGLGSKKRDRLLLRDFNEKDVDFRERNINDT RYMARYIMKYLDDCIDFSQSKADVKDRVQARNGALTDFLRHQWGLKKNRSESDKHHAQDAIVIACATNGYTQYL AHLSKIFENKQAYANKYGEPWYKAFKQHVKQPWDGFYQDVQASLAEIFVSRPPRKNATGEIHQDTVHTLNPKKK NRRGELVYSAKDVKSGLNVRGGMAGNSNMFRCDVFEKGGRFFIVPVYLADFCEKNRSDCFCPKDKEHEKPDE TYKFLFSVFKDDYLSIETNKGEKFYGYMNQYITSTGQFYIGSNDSSPMYSISTSSFEKQDILLYRIGEKFEKCSVV DFDSEKQEMKVASLEDGSIHKIKASAKVNKKGEVQKIYKTDKPYEKLDKGKTISSATFKRIRKYQVDPLGRYVEVK SEIRLPLNIKKGKA
Al3Cas9	MSLIFGFDLGIASIGWAAVKMSDPTASDTDAAAPDNRPVGEILGAGVRCFDQAQNAVDRRVARGARRRIRHRAQ RMRDIRKILRQYNLIDIPEPKFNGQNNFYLNRDTDINIWQLRAVDAFTRKLTPRETGRILYHLAKHRGYDDITYPVL TGAVPEKADDAEAKEELKAIGAIKDNFAKLHNENPDQTMCQMLYATEHGQMRNGRKSVIKITKDGTQRTEEQSS YSNSIPRSEIKREAEMILHAQCGFGNDFSDVFKKWCVVAFHQLKFNESKSPLAKSIEKMRGKCPFTNKPVAPKES PTAQMFVALTKCRQNGFTVAETNAIIDALYKKKTGLKYADVRKLLKWDDDSQFKTLNYSRTYDKENKKWIEPDVK KIENSKFYAFTGWHKLAKYTTDVATMDKIFDIIVTQKTPENIERAVSGILPEHAAAIAQITSSEFIKLSISALQKIIPEM AAGKKYNEAVADGLGKDFRVQNVSYIDRDNGVSAGCLHQINWNALGNRITSPVAKRTLGQLRKVYNAMVYRFG APDRIHLEIGRELKNSPDDIARMNREQERNRQATEEARKEYGKNYFKYKLYMEQGGKCPYCGATLVAGDWAAY EIDHILPYSRSLDNSQSNKVLVCHECNQRKGNKTPYEFLTTEQFHDMAVRARALHNPAKFKKLTNTELPQSSDE QNGFIERNANDNATIARFATEYLEYGIDWPESDNIKKRVLVRTGSLTDYLRHQWGLVKNRDESDKHHAQDAIVIA CATQGMVKYLSTLSARFENKYTLIEKYGQAWYNVLKTSVQEPWAGFRTAVFSLLDNIIVSRPPRCNATGAAHKET IYANPKSYRSKRGLNKNRDKGSMKIRHGRVERGDMFRFDIWRDARGKYMCVPIFVADTKGKDDNNFKVPDSEF ICSLHKDDYVRLTTADAVYEGYITQMDHGTNFILYRQDNSQIQVVKKSVGKCTNIEKFAMSLLGTPKPIKLPEQRT PVVK
CaCas9	MEITINREIGKLGLPRHLVLGMDPGIASCGFALIDTANHEILDLGVRLFDSPTHPKTGQSLAVIRRGFRSTRRNIDR TQARLKHCLQVLKAYGLIPQDATKEYLHTTKGDKQPLKLRVDGLDRLLNDREWALVLYSLCKRRGYIPHGEGNQ DKSSEGGKVLSALAANKEAIAETSCRTVGEWLAWQPQSRNRGGNYDKCVTHAQLIEETHILFDAQRSFGSKYA SPEFEAAYIEVCDWERSRKDFDRRTYDLVGHCSYFPTEKRAARCTLTSELVSAYGALGNITIIHENGTSRALSATE RDECIAILFSCEPIRGNKDCAVKFGALRKALDLSSGDYFKGVPAADEKTREVYKPKGWRVLRNTLNAANPILLQR LRDDRNLADAVMEAVAYSSALPVLQEQLQGLPFSEAEIEALCRLPYSSKALNGYGNRSKKALDMLLDCLEEPEVL NLTQAENDCGLLGLRIAGAQLERSDRLMPYETWIELTGRTNNNPVVIRSMSQMRKVVNAVCRKWGVPNEIHVEL DRELRLPQRAKDEIAKANKKNEKNRERIAGQIAELRGCTADEVTGKQIEKYRLWEEQECFDLYTGAKIEVDRLISD DTYTQIDHILPFSRTGENSRNNKVLVLAKSNQDKREQTPYEWMSHDGAPSWDAFERRVQENQKLSRRKKNFLL EKDLDTKEGEFLARSFTDTAYMSREACAYLADCLLFPDDGAKAHVVPTTGRATAWLRRRWGLNFGSNGEKDRS DDRHHATDACVIAACSRSLVIKTARINQETHWSITRGMNETQRRDAIMKALESVMPWETFANEVRAAHDFVVPT RFVPRKGKGELFEQTVYRYAGVNAQGKDIARKASSDKDIVMGNAVVSLDEKSVIKVSEMLCLRLWHDPEAKKG QGAWYADPVYKADIPALKDGTYVPRIAKAHTGRKAWKPVPESAMKKPPLEIYLGDLVQIGDFMGRFSGYNIANA NWSFVDRLTKEALGCPTVVKLDNKLAPAIIRESIIMH
TmCas9	QPESVRLFGKLGVRRYLVLGLDPGVASCGFSLIDLHNHEILEMGSRLFDSPVVPKTQQSKAAVRRGFRSTRRNL DRTQNRLKHCLRLLKEQGLVPAGATPEFFHTAKGDQPPIRLRVAGLDRLLTDREWAIVLYSLCKRRGYIPHGEGG EGDVDSASEDGKVLKAISQNKKLLAESGARTVGEWLATQPRSRNRSGEYDKCVTHAQLTDEVRALFSAQRSFG STSATEDLEDAYLEVFDWEKPRDAFDRRSYELVGNCVYFPGEKRAARCTLTSELVSAYGALGNVTIQFPDGTTR PLTAKERDTFIDTLFSPSALKGNKECKVKFSDIRRRLDLDARTSFKGISSDDEKNREVYRPQGWRTLRKTLGSEG TALLTRFLADRDLADATMEAVAYSSSLPVLSSRLEELELTDDEVALVERLPYSSRALNGYGSRSKKALDLLLGAFE EPEVLTLTDAETATGLDELRKHGSAGIERSTKLMPYAAWIALTGRTNNNPVVLRAMAQMRKVVNAVCGEWGVPN EIHVELARELALPKKAKDSIARANRQNEKDNERIRKQISELTGRDHDRIKNSLVAKWRLWEEQGNCDIYTGERIDA LRLVNDDTYTQVDHILPFSRTGDNSRHNKVLVLAHSNQLKRERSPFEWMTSEEAGAPDWNSFVASVQENRRIS PRKRGFLLERDLMSKEGDFQSRNLTDTAYMAREVCAYLSDCLAFPNDGRKVHVVPTKGTASAWLRRSWGLNF GMAGEKDRSDDRHHATDACVIAACSRSLVIQTAKLSEHHYTLNEQGRDEALLGSMPWPSFADDVRQMREKVIP TRFVPRVGSGELFEQTIYSYVGTDDKGKDLLSAKGGKPKPAGNAVVSDDEKSAIKVGGMICLRLWHDVDARKG RGQWYADPIYYADLPALRNGTYVPRIAQAHKGRKTWKPIPERVLSGKPICLYLGDAVRVDHEIGRFSGFDIDAAN WSLKSIYSGEKICFPSIGKIDNITFPIKITEDILGRCWGNL

Table 2. Sequences of the primers used for NGS analysis (PAM identification assay).

Primer name	**Sequence (5' → 3')**
F1a_library	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAAAACACACCGCATACGTACGATTTA
F1b_library	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTTAATCACACCGCATACGTACGATTTA
R1_library	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCGTTCTGATTTAATCTGTATCAGGC
F2a_library	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTTTTGGTTACGCATCTGTGCGGTATTTC
F2b_library	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCCTTACGCATCTGTGCGGTATTTC
F3a_library	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTATCGATTTAAATAGGCCTGACTCAC
F3b_library	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATTTCGATTTAAATAGGCCTGACTCACTA
R2_library	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAATCTTCTCTCATCCGCCAAA
F4a_invitro	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTGCTGAACCGCTCTTCCGATC
F4b_invitro	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTAAGACTGCTGAACCGCTCTTCCGATC
F4c_invitro	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCTAGACCTAATGTGATCTGCTGAACCGCTCTTCCGATC
R3_invitro	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCTGCGTTCTGATTTAATCTGTATCAGGC
Oligo UP dsDNA adapter	CGGCATTCCTGCTGAACCGCTCTTCCGATCT
Oligo BOTTOM dsDNA adapter	GATCGGAAGAGCGGTTCAGCAGGAATGCCG
Sequences of the primers used for NGS analysis (deep sequencing).	
Primer name	**Sequence (5' → 3')**
NGS_RHO_Fw	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGgggtcagccacaagggc
NGS_RHO_Rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCCAGCATGGAGAACTGC

Table 3. Parameters employed for the estimation of base editors' editing window.

Variant	**Spacer length**	**PAMs**	**Window (>10% editing)**
CoABE8e	23	NNNNATBT, NNNNGCBT, NNNNGWNT	6, 8-10, 12-14, 16-17, 19
SaABE8e	21	NNGRRT	8-16, 18-19
CjABE8e	22	NNNVRYAC	9, 11, 16
Nme2ABE8e	22	NNNNCC	11-14, 16-19
SauriABE8e	21	NNGG	7-17, 19

Table 4. Oligonucleotides used to amplify DNA templates to perform TIDE analysis, *in vitro* cleavage and qPCR.

Oligos to amplify genomic regions for TIDE analysis		
Locus	For (5' → 3')	Rev (5' → 3')
EMX1_1	ATTTCGGACTACCCTGAGGAG	GGAATCTACCACCCCAGGCTCT
EMX1_2	GAAGCGATTATGATCTCTCC	GGAATCTACCACCCCAGGCTCT
FAS_1	TTAGAAAGGGCAGGAGGC	CTTGTCCAGGAGTTCCGCTC
FAS_2	AATTGAAGCGGAAGTCTGGG	AACACTTCTCTCGCTATGCC
CCR5	ATGCACAGGGTGGAACAAGATGGA	CTAAGCCATGTGCACAACTCTGAC
FANCF	GGCACATCTTGGGACTCAG	AGCATAGCGCCTGGCATTAATAGG
HBB	CAAAGAACCTCTGGGTCCAAG	GCATATTCTGGAGACGCAGG
ZSCAN2	GACTGTGGGCAGAGGTTCAGC	TGTATACGGGACTTGACTCAGACC
CHR6	ATGTCCTCATGCCGGACTG	TCCAAGAGCATACGCACACATTCC
ADAMTSL1	TAGGACTAGGCTCTTGGAG	CATAGAGTACTTAGTATGAGCGAGGC
B2M	CCAGTCTAGTGCATGCCTTC	GTTCCCATCACATGTCAC
CXCR4	GGACAGGATGACAATACCAGGCAGGATAAGGCC	AGAGGAGTTAGCCAAGATGTGACTTTGAAACC
PD1_1	ACGTCGTAAAGCCAAGGTTAGTCC	CACCCTCCCTTCAACCTGACC
PD1_2	TCCAGTTCTGAGTCCTGAGTG	GGACTAACCTTGGCTTTACGACG
DNMT1	GTCTTAATTTCCACTCATACAGTGGTAG	CGTTTTGGGCTCTGGGACTCAG
MATCH8	TGTGTCGTCCATAAACGCTGCC	CATCTTCCCTGAAATTTCTTAAGAGGC
TRAC	CTGTCCCTGAGTCCCAGT	GGCCTAGAAGAGCAGTAAGG
TRBC	CTGACCACGTGGAGCTGAG	CTTACTTACCCGAGGTAAAGCC
VEGFAsite2	TGCGAGCAGCGAAAGCGACA	TCCAATGCACCCAAGACAGC
VEGFAsite3	GCATACGTGGGCTCCAACAGGT	CCGCAATGAAGGGGAAGCTCGA
CACNA2D4	TACAGCAGGACTGTGTGGCACG	CTTCCATCCTCCATCAGGTCAGG
HEKsite1	GAAGGATAGAGGGTGGGAGAGG	TGGAGTGCAATGGCGTGAC
HEKsite1_2	TCAACAGGGCTTCGTTACCC	TGGAGTGCAATGGCGTGAC
HEKsite3	TAGCTACGCCTGTGATGG	CCAGAGAAGTTGCTAGGATGAAAGG

HEKsite4	AACAATTTCAGATCGCGG	GTCAGACGTCCAAAACCAGACTCC
CHR8	TCCTGGGTCTGAGTTTCTGAGAGG	ACAACACAGATCTGCAGATCTCCG
BCR	GTCAGGGCGCTCCTTCCTTC	GTGTACAGGGCACCTGCA
ATM	CTAAGGGGTCTGACACAGACTG	GTGGCTACAAGACATTTCCTCC
HBG1	GCCTGTGAGATTGACAAGAACAG	TACTGCGCTGAAACTGTGGC
HPRT_1	ACAGTTACTAATATCATCTTACACC	GGCTGAAAGGAGAGAACT
HPRT_2	CAGCAGCTGTTCTGAGTACTTG	CCCTTGACCCAGAAATTCCAC
IL2RG	CTGGTTTGGATTAGATCAGAGG	GTTCCAAGTGCAATTCATG
NF1	GCAGTACTGCAAGCATCCTG	GCTCCAAGATGGCCAACTAGC
USH2A_1	TCCACATCCCTCCCTTTCATG	CCAGAGTAGAAGGCAGCTAGC
USH2A_2	GCTACTGCTGTGAGCATAACT	TGGGGATACAGCTCCTTTCT
RHO	CTGATTCAGCCAGGAGCTTAGG	GAGATAGATGCGGGCTTCCA
BCLenh	GGACTTGGGAGTTATCTGTAG	GAGGCAAGTCAGTTGGGAAC
BCLenh_2	GCTAAGAATAGTGAAAACACCCTTG	GTTCCCAACTGACTTGCCTC
BCLenh_3	GCTAAGAATAGTGAAAACACCCTTG	GTCTGGCTTTGCCAATTGGAG
HEKsite2	CACTGCCATTCTACCAACAATAGAGG	CTAAAACATCCAACCTTGATAGAACAC C
HEKsite2_2	CACTGCCATTCTACCAACAATAGAGG	CCCTTCAAATAAGTAGCTGCTTTGGAT C
AAVS1	TCTCCTCTTGGGAAGTGTAAGG	CCTGTGCCATCTCTCGTTTC
PCSK9_1	GCCAGCTGATGGCCTTGG	CCAGCAAGTGTGACAGTCATG
PCSK9_2	GCACAGATGCCATCTGGTTG	GTCCTCACAAGTGTGATCCATG

Oligos to amplify DNA templates for in vitro cleavage experiments	
Oligo name	**5’ → 3’**
oligo_lmw-E (EGFP-F)	AAATGGGCGGTAGGCGTG
oligo_hmw-E (EGFP-R)	AGCTCGTCCATGCCGAGAGTGATC

oligo_lmw-B (B2M-F)	CTCTAACCTGGCACTGCGTC
oligo_hmw-B (B2M-R)	CGAAGTCCACAGCTCTCCAG
Oligos for in vitro transcription	
Target	**5' → 3'**
CoCas9-EGFP-sgRNA-IVT_F	TAATACGACTCACTATAGGcctcgccggacacgctgaacttgtGTCTTGAGCACGCGAAAGCG
CoCas9-B2M-sgRNA-IVT_F	TAATACGACTCACTATAGGCTGGATAGCCTCCAGGCCAGAAAGTCTTGAGCACGCGAAAGCG
CoCas9-sgRNA-IVT_R	AAAAAAGTTCGCGCCGCCCA

Table 5. Sequences of the oligonucleotides used for cloning sgRNA spacers and sequences of their relative target sites.

Spacer sequences used in reporter assays	**Oligo used to clone the spacer in pX plasmid (##)**	
Name (*)	**Oligo 1 (5'>3')**	**Oligo 2 (5'>3')**
gRNA_CoCas9_EGFP	caccGCCTCGCCGGACACGCTGAACTTGT	agacACAAGTTCAGCGTGTCCGGCGAGGC
gRNA_SpCas9_EGFP	caccGGGCACGGGCAGCTTGCCGG	gaacCCGGCAAGCTGCCCGTGCCC
gRNA_CaCas9_EGFP	caccGACGGCAACTACAAGACCCGCGC	agacGCGCGGGTCTTGTAGTTGCCGTC
gRNA_AI1Cas9_EGFP	caccGCCCATCCTGGTCGAGCTGGACGG	taacCCGTCCAGCTCGACCAGGATGGGC
gRNA_GeCas9_EGFP	caccGGCCGACAAGCAGAAGAACGGCAT	tgacATGCCGTTCTTCTGCTTGTCGGCC
gRNA_AI3Cas9_EGFP	caccGACTTCTTCAAGTCCGCCATGCCC	taacGGGCATGGCGGACTTGAAGAAGTC
gRNA_ClCas9_EGFP	caccGCTTGTGCCCCAGGATGTTGCCGT	caatACGGCAACATCCTGGGGCACAAGC
gRNA_SuCas9_EGFP	caccGCCAGGGCACGGGCAGCTTGCCGG	gaacCCGGCAAGCTGCCCGTGCCCTGGC
gRNA_BsCas9_EGFP	caccGCGTGCAGTGCTTCAGCCGCTACC	tgatGGTAGCGGCTGAAGCACTGCACGC
gRNA_AI4Cas9_EGFP	caccGATCCTGGTCGAGCTGGACGGCG	taacCGCCGTCCAGCTCGACCAGGATC

gRNA_Al2Cas9_EGFP	caccGCAGCTCGACCAGGATGGGCACCAC	taacGTGGTGCCCATCCTGGTCGAGCTGC
gRNA_TmCas9_EGFP	caccGACGGCGTGCAGTGCTTCAGCCGC	caccGACGGCGTGCAGTGCTTCAGCCGC
Spacer sequences of in vitro transcribed sgRNA		
gRNA_CoCas9_EGFP	CCTCGCCGGACACGCTGAACTTGT	cgcCCTCGCCGGACACGCTGAACTTGTGGCCGTTTacg
gRNA_CoCas9_B2M-IVC	CTGGATAGCCTCCAGGCCAGAAA	acgCTGGATAGCCTCCAGGCCAGAAAGAGAGAGTagc
Spacer sequences to target endogenous loci	**Oligo used to clone the spacer in pX plasmid (##)**	
Name (*) (**)	**Oligo 1 (5'>3')**	**Oligo 2 (5'>3')**
gRNA1_CoCas9_EMX1	caccGTGCCCCTCCCTCCCTGGCCCAGGT	agacACCTGGGCCAGGGAGGGAGGGGCAC
gRNA2_CoCas9_EMX1	caccGGGCATGGTTTCATAACTAGGAGG	agacCCTCCTAGTTATGAAACCATGCCC
gRNA1_CoCas9_FAS	caccGCGCCTGGGCAGCCAGGGCTGGCCT	agacAGGCCAGCCCTGGCTGCCCAGGCGC
gRNA2_24_CoCas9_FAS	caccGCCTGGGCAGCCAGGGCTGGCCTCA	agacTGAGGCCAGCCCTGGCTGCCCAGGC
gRNA2_23_CoCas9_FAS	caccGCTGGGCAGCCAGGGCTGGCCTCA	agacTGAGGCCAGCCCTGGCTGCCCAGC
gRNA2_22_CoCas9_FAS	caccGTGGGCAGCCAGGGCTGGCCTCA	agacTGAGGCCAGCCCTGGCTGCCCAC
gRNA1_CoCas9_CCR5	caccGCAAGAGGCTCCCGAGCGAGCAAG	agacCTTGCTCGCTCGGGAGCCTCTTGC
gRNA1_CoCas9_FANCF	caccGGAAGGCCGAAGCGGAGCGTCCCG	agacCGGGACGCTCCGCTTCGGCCTTCC
gRNA2.1_CoCas9_FANCF	caccGCGCGCTACCTGCGCCACATCCAT	agacATGGATGTGGCGCAGGTAGCGCGC
gRNA2.2_CoCas9_FANCF	caccGGGCAAGGCGGGCCAGGCTCTCT	agacAGAGAGCCTGGCCCGCCTTGCCC
gRNA1_CoCas9_HBB	caccGAGATATATCTTAGAGGGAGGGCT	agacAGCCCTCCCTCTAAGATATATCTC
gRNA2_24_CoCas9_HBB	caccGTCTCCTCAGGAGTCAGGTGCACCA	agacTGGTGCACCTGACTCCTGAGGAGAC
gRNA2_23_CoCas9_HBB	caccGCTCCTCAGGAGTCAGGTGCACCA	agacTGGTGCACCTGACTCCTGAGGAGC
gRNA2_22_CoCas9_HBB	caccGTCCTCAGGAGTCAGGTGCACCA	agacTGGTGCACCTGACTCCTGAGGAC
gRNA1_CoCas9_ZSCAN2	caccGCATGAGGCATTTGTAGGGCTTCTC	agacGAGAAGCCCTACAAATGCCTCATGC

gRNA2_CoCas9_ZSCAN2	caccGGCTTCTCCACCATGTGGGTTCTC	agacGAGAACCCACATGGTGGAGAAGCC
gRNA1_CoCas9_Chr6	caccGTTGATTCTTACAACAACATGAGAG	agacCTCTCATGTTGTTGTAAGAATCAAC
gRNA1_CoCas9_ADAM	caccGGGGCAGAGAGAGAGAGTGAGCGA	agacTCGCTCACTCTCTCTCTCTGCCCC
gRNA2_CoCas9_ADAM	caccGAAAGAAATACTAAGACATGCAGAG	agacCTCTGCATGTCTTAGTATTTCTTTC
gRNA1_CoCas9_B2M	caccGGGCCTTGTCCTGATTGGCTGGGC	agacGCCCAGCCAATCAGGACAAGGCCC
gRNA2_CoCas9_B2M	caccGCTGAGGTTTGTGAACGCGTGGAG	agacCTCCACGCGTTCACAAACCTCAGC
gRNA1_CoCas9_CXCR4	caccGTCATCTACACAGTCAACCTCTAC	agacGTAGAGGTTGACTGTGTAGATGAC
gRNA2_CoCas9_CXCR4	caccGTTGGCTGAAAAGGTGGTCTATGT	agacACATAGACCACCTTTTCAGCCAAC
gRNA2.1_CoCas9_PD1	caccGTCGGCGGTCAGGTGTCCCAGAGCC	agacGGCTCTGGGACACCTGACCGCCGAC
gRNA2.2_CoCas9_PD1	caccGTCCCTCTGGGCCCTGCTGGGTAC	agacGTACCCAGCAGGGCCCAGAGGGAC
gRNA1_CoCas9_DNMT1	caccGTCGCCTGTCAAGTGGCGTGACACC	agacGGTGTCACGCCACTTGACAGGCGAC
gRNA3.1_CoCas9_DNMT1	caccGAAAAGTCACTCTGGGGAACACG	agacCGTGTTCCCCAGAGTGACTTTTC
gRNA3.2_CoCas9_DNMT1	caccGTCAGCTGTTAACATCAGTACGT	agacACGTACTGATGTTAACAGCTGAC
gRNA3.3_CoCas9_DNMT1	caccGACTAGTTCTGCCCTCCCGTCAC	agacGTGACGGGAGGGCAGAACTAGTC
gRNA1_CoCas9_Match8	caccGGGAGGTGGCAGGGGGAGGAAAGC	agacGCTTTCCTCCCCCTGCCACCTCCC
gRNA1_CoCas9_TRAC (#) (TRAC-g1)	caccGATAAGGCCGAGACCACCAATCAG	agacCTGATTGGTGGTCTCGGCCTTATC
gRNA3.1_CoCas9_TRAC	caccGAGGCAGACAGACTTGTCACTGGA	agacTCCAGTGACAAGTCTGTCTGCCTC
gRNA3.2_CoCas9_TRAC	caccGAATGTGTCACAAAGTAAGGATTC	agacGAATCCTTACTTTGTGACACATTC
gRNA3.3_CoCas9_TRAC (#) (TRAC-g3.3)	caccGGCTCACTGTTTCTTAGTAAAAA	agacTTTTTACTAAGAAACAGTGAGCC
gRNA1_CoCas9_TRBC	caccGCCTCGGCGCTGACGATCTGGGTG	agacCACCCAGATCGTCAGCGCCGAGGC
gRNA2_CoCas9_TRBC	caccGTCAGAGGAAGCTGGTCTGGGCCT	agacAGGCCCAGACCAGCTTCCTCTGAC
gRNA1_CoCas9_VEGFAsite2	caccGAGGAGGTGGTAGCTGGGGCTGGG	agacCCCAGCCCCAGCTACCACCTCCTC
gRNA2_CoCas9_VEGFAsite2	caccGGAGGTGGTAGCTGGGGCTGGGGG	agacCCCCCAGCCCCAGCTACCACCTCC

gRNA2_CoCas9_VEGFAsite3	caccGCCCATTCCCTCTTTAGCCAGAG C	agacGCTCTGGCTAAAGAGGGAATGGG C
gRNA1_CoCas9_CACNA2D4	caccGAGAGAGGCTCCCATCACGGGG GA	agacTCCCCCGTGATGGGAGCCTCTCT C
gRNA1_CoCas9_HEKsite3 (#) (H3-g1)	caccGCAGCAGAAATAGACTAATTGCAT	agacATGCAATTAGTCTATTTCTGCTGC
gRNA1_CoCas9_HEKsite4	caccGCCAGGTCAGATAAATTTTAGGAA G	agacCTTCCTAAAATTTATCTGACCTGGC
gRNA1_CoCas9_Chr8	caccGACCCCTAATATGAAGGGTAGAG T	agacACTCTACCCTTCATATTAGGGGTC
gRNA2_CoCas9_Chr8	caccGGCAGGGCCTGACAGCGGAAAG GG	agacCCCTTTCCGCTGTCAGGCCCTGC C
gRNA1_CoCas9_BCR	caccGTGGCTGTGCTTAGGTAGCGTGG G	agacCCCACGCTACCTAAGCACAGCCA C
gRNA2_CoCas9_BCR	caccGAGTTCTTGCCGTGCCCCTTCCC CA	agacTGGGGAAGGGGCACGGCAAGAA CTC
gRNA3.1_CoCas9_BCR	caccGCCCGTCCTCTGCTAACACAGGC	agacGCCTGTGTTAGCAGAGGACGGGC
gRNA3.2_CoCas9_BCR	caccGCGTGGAGCTGCAGATGCTGAC CA	agacTGGTCAGCATCTGCAGCTCCACG C
gRNA3.3_CoCas9_BCR	caccGCTGGACTGGGAGCACCTCCCT G	agacCAGGGAGGTGCTCCCAGTCCAGC
gRNA1_CoCas9_HEKsite1 (#) (H1-g1)	caccGAGAGCTCTTGGTACCTGAAGTT AT	agacATAACTTCAGGTACCAAGAGCTCT C
gRNA3.1_CoCas9_HEKsite1 (#) (H1-g3.1)	caccGCCCCTTGTGGTGAGGGAGAGC CG	agacCGGCTCTCCCTCACCACAAGGGG C
gRNA3.2_CoCas9_HEKsite1	caccGCAGCGTCTTCCCTTCCTCCAGC G	agacCGCTGGAGGAAGGGAAGACGCTG C
gRNA3.3_CoCas9_HEKsite1	caccGTTTAGTTCTCTTCCTCTTTCCCC	agacGGGGAAAGAGGAAGAGAACTAAA C
gRNA1_CoCas9_HBG1	caccGAGGATACCAGGACTTCTTTTGT C	agacGACAAAAGAAGTCCTGGTATCCTC
gRNA2_CoCas9_HBG1	caccGAAATGACCCATGGCGTCTGGAC T	agacAGTCCAGACGCCATGGGTCATTTC
gRNA1_CoCas9_HPRT	caccGCTTGCATTGTATGTCTGGCTATT	agacAATAGCCAGACATACAATGCAAGC
gRNA2_CoCas9_HPRT	caccGATCATTATGCTGAGGATTTGGAA A	agacTTTCCAAATCCTCAGCATAATGATC
gRNA1_CoCas9_IL2RG	caccGCACCTAATCTCCTAGAGGACTTA	agacTAAGTCCTCTAGGAGATTAGGTGC
gRNA2_CoCas9_IL2RG	caccGAGGCAGGAGGATCACTAGAGG CC	agacGGCCTCTAGTGATCCTCCTGCCTC
gRNA2_CoCas9_ATM	caccGTTAGGAACTTTATTGGCTGGAAC	agacGTTCCAGCCAATAAAGTTCCTAAC
gRNA2_CoCas9_NF1	caccGTTGAGGGGTCTCCTTGTGGTGG	agacCCACCACAAGGAGACCCCTCAAC

gRNA3.1_CoCas9_USH2A	caccGACTAAATTGAAAGAGTGCCAGGAG	agacCTCCTGGCACTCTTTCAATTTAGTC
gRNA3.2_CoCas9_USH2A	caccGTGATTCTGGAGAGGAAGCTGAA	agacTTCAGCTTCCTCTCCAGAATCAC
gRNA3.3_CoCas9_USH2A	caccGCTTTAAGAAAAGGCTGTGTATTGT	agacACAATACACAGCCTTTTCTTAAAGC
gRNA1_CoCas9_BCLenh (#)	caccGGAATGAAATAATTTGTATGCCA	agacTGGCATACAAATTATTTCATTCC
gRNA1_CoCas9_HEKsite2 (#)	caccGCCCTGTAAAGGAAACTGGAACA	agacTGTTCCAGTTTCCTTTACAGGGC
gRNA2_CoCas9_AAVS1	caccGGTGACCCGAATCCACAGGAGAA	agacTTCTCCTGTGGATTCGGGTCACC
gRNA3.1_CoCas9_RHO	caccGTTCTACGTGCCCTTCTCCAATGCG	agacCGCATTGGAGAAGGGCACGTAGAAC
gRNA3.2_CoCas9_RHO	caccGGTAGTACTGTGGGTACTCGAAGG	agacCCTTCGAGTACCCACAGTACTACC
gRNA3.3_CoCas9_RHO	caccGTTCGGGCCCACAGGATGCAATTTG	agacCAAATTGCATCCTGTGGGCCCGAAC
BEgRNA2_CoCas9_BCLenh (BCL-BEg2)	caccGAAACCACAGGGATCACAACACAT	agacATGTGTTGTGATCCCTGTGGTTTC
BEgRNA3_CoCas9_BCLenh (BCL-BEg3a)	caccGGTGTGCATAAGTAAGAGCAGAT	agacATCTGCTCTTACTTATGCACACC
BEgRNA3_CoCas9_BCLenh (BCL-BEg3)	caccGTCCAGTGCAAAGTCCATACAGGT	agacACCTGTATGGACTTTGCACTGGAC
BEgRNA4_CoCas9_BCLenh (BCL-BEg4)	caccGAATAATAATAGTATATGCTTCA	agacTGAAGCATATACTATTATTATTC
BEgRNA1a_CoCas9_HEKsite1 (H1-BEg1a)	caccGCTGCATTGACTCAGGCCATTCTA	agacTAGAATGGCCTGAGTCAATGCAGC
BEgRNA1b_CoCas9_HEKsite1 (H1-BEg1b)	caccGCCCACCTCGGCCTCTCAAAGTG	agacCACTTTGAGAGGCCGAGGTGGGC
BEgRNA1_CoCas9_HEKsite1 (H1-BEg1)	caccGGAGGAGCACAAAGTGGGAAAGA	agacTCTTTCCCACTTTGTGCTCCTCC
BEgRNA2_CoCas9_HEKsite1 (H1-BEg2)	caccGATCAATAGCTATGGTCAGATACG	agacCGTATCTGACCATAGCTATTGATC
BEgRNA2_CoCas9_HEKsite2 (H2-BEg2)	caccGAAAAATGATATCCATTATTAGT	agacACTAATAATGGATATCATTTTTC
BEgRNA3_CoCas9_HEKsite2 (H2-BEg3a)	caccGAACATGCCCATATGTGGATCTGC	agacGCAGATCCACATATGGGCATGTTC
BEgRNA3_CoCas9_HEKsite2 (H2-BEg3)	caccGCACTATGTGTCACTGTGTGTAA	agacTTACACACAGTGACACATAGTGC
BEgRNA4_CoCas9_HEKsite2 (H2-BEg4)	caccGATGCTGAGGGGAGCAGTGGGGGA	agacTCCCCCACTGCTCCCCTCAGCATC
BEgRNA1_CoCas9_HEKsite3 (H3-BEg1)	caccGATGGAGCTTGCCCAGGACCCCA	agacTGGGGTCCTGGGCAAGCTCCATC

BEgRNA2_CoCas9_HEKsite3 (H3-BEg2)	caccGCAGCAGAAATAGACTAATTGCAT	agacATGCAATTAGTCTATTTCTGCTGC
BEgRNA1_CoCas9_HEKsite4 (H4-BEg1)	caccGCAGGGGAGGGGAAAGAAGACCA	agacTGGTCTTCTTTCCCCTCCCCTGC
BEgRNA2_CoCas9_HEKsite4 (H4-BEg2)	caccGAGTCACCATCACAAGGAAACGCT	agacAGCGTTTCCTTGTGATGGTGACTC
BEgRNA1_CoCas9_PCSK9 (PCSK-BEg1)	caccGAAAGACGGAGGCAGCCTGGTGG	agacCCACCAGGCTGCCTCCGTCTTTC
BEgRNA2_CoCas9_PCSK9 (PCSK-BEg2)	caccGCCCACAACGCTTTTGGGGGTGA	agacTCACCCCCAAAAGCGTTGTGGGC
gRNA2_SpCas9_EMX1	caccGTGGTTTCATAACTAGGAGG	aaacCCTCCTAGTTATGAAACCAC
gRNA2_SpCas9_FAS	caccGGCAGCCAGGGCTGGCCTCA	aaacTGAGGCCAGCCCTGGCTGCC
gRNA2_SpCas9_FANCF	caccGCTACCTGCGCCACATCCAT	aaacATGGATGTGGCGCAGGTAGC
gRNA2_SpCas9_HBB	caccGTCAGGAGTCAGGTGCACCA	aaacTGGTGCACCTGACTCCTGAC
gRNA2_SpCas9_ZSCAN2	caccGCTCCACCATGTGGGTTCTC	aaacGAGAACCCACATGGTGGAGC
gRNA2_SpCas9_CHR6	caccGTCTTACAACAACATGAGAG	aaacCTCTCATGTTGTTGTAAGAC
gRNA2_SpCas9_ADAM	caccGAATACTAAGACATGCAGAG	aaacCTCTGCATGTCTTAGTATTC
gRNA2_SpCas9_B2M	caccGGGTTTGTGAACGCGTGGAG	aaacCTCCACGCGTTCACAAACCC
gRNA2_SpCas9_CXCR4	caccGCTGAAAAGGTGGTCTATGT	aaacACATAGACCACCTTTTCAGC
gRNA2_SpCas9_PD1	caccGGGTCAGGTGTCCCAGAGCC	aaacGGCTCTGGGACACCTGACCC
gRNA2_SpCas9_DNMT1	caccGTGTCAAGTGGCGTGACACC	aaacGGTGTCACGCCACTTGACAC
gRNA2_SpCas9_TRAC (#)	caccGGGCCGAGACCACCAATCAG	aaacCTGATTGGTGGTCTCGGCCC
gRNA2_SpCas9_TRBC	caccGAGGAAGCTGGTCTGGGCCT	aaacAGGCCCAGACCAGCTTCCTC
gRNA2_SpCas9_VEGFAsite2	caccGTGGTAGCTGGGGCTGGGGG	aaacCCCCCAGCCCCAGCTACCAC
gRNA2_SpCas9_VEGFAsite3	caccGTTCCCTCTTTAGCCAGAGC	aaacGCTCTGGCTAAAGAGGGAAC
gRNA2_SpCas9_CACNAD24	caccGAGGCTCCCATCACGGGGGA	aaacTCCCCCGTGATGGGAGCCTC
gRNA2_SpCas9_HEKsite3 (#)	caccGAGAAATAGACTAATTGCAT	aaacATGCAATTAGTCTATTTCTC
gRNA2_SpCas9_HEKsite4 (#)	caccGACCATCACAAGGAAACGCT	aaacAGCGTTTCCTTGTGATGGTC
gRNA2_SpCas9_Chr8	caccGGGCCTGACAGCGGAAAGGG	aaacCCCTTTCCGCTGTCAGGCCC
gRNA2_SpCas9_BCR	caccGTTGCCGTGCCCCTTCCCCA	AAACTGGGGAAGGGGCACGGCAAC
gRNA2_SpCas9_HEKsite1 (#)	caccGTCTTGGTACCTGAAGTTAT	aaacATAACTTCAGGTACCAAGAC
gRNA2_SpCas9_HBG1	caccGGACCCATGGCGTCTGGACT	aaacAGTCCAGACGCCATGGGTCC
gRNA2_SpCas9_HPRT	caccGTATGCTGAGGATTTGGAAA	aaacTTTCCAAATCCTCAGCATAC
gRNA2_SpCas9_IL2RG	caccGAGGAGGATCACTAGAGGCC	aaacGGCCTCTAGTGATCCTCCTC

BEgRNA1_SpCas9_BCLenh	caccGTTATCACAGGCTCCAGGAA	aaacTTCCTGGAGCCTGTGATAAC
BEgRNA1_SpCas9_HEKsite2	caccGAACACAAAGCATAGACTGC	aaacGCAGTCTATGCTTTGTGTTC
BEgRNA1_SpCas9_Pcsk9	caccGGACGGAGGCAGCCTGGTGG	aaacCCACCAGGCTGCCTCCGTCC
BEgRNA2_SpCas9_Pcsk9	caccGACAACGCTTTTGGGGGTGA	aaacTCACCCCCAAAAGCGTTGTC
gRNA2_Nme2Cas9_ATM	caccGAACTTTATTGGCTGGAACTGG	caacCCAGTTCCAGCCAATAAAGTTC
gRNA2_Nme2Cas9_CHR6	caccGCTACTTGTGAAACATCTTGAAC	caacGTTCAAGATGTTTCACAAGTAGC
gRNA2_Nme2Cas9_NF1	caccGAGGGGTCTCCTTGTGGTGGAGG	caacCCTCCACCACAAGGAGACCCCTC
gRNA2_Nme2Cas9_DNMT1	caccGTTACTCGCCTGTCAAGTGGCGT	caacACGCCACTTGACAGGCGAGTAAC
gRNA3.1_Nme2Cas9_DNMT1	caccGCACAACATCAGTGCATGTTGGG	caacCCCAACATGCACTGATGTTGTGC
gRNA3.2_Nme2Cas9_DNMT1	caccGAGACTGAACACTCCTCAAACG	caacCGTTTGAGGAGTGTTCAGTCTC
gRNA3.3_Nme2Cas9_DNMT1	caccGCTCGGTGAATTTGGCTCAGCA	caacTGCTGAGCCAAATTCACCGAGC
gRNA2_Nme2Cas9_USH2A	caccGTTTTACTATTCAACAATTAGG	caacCCTAATTGTTGAATAGTAAAAC
gRNA3.1_Nme2Cas9_USH2A	caccGCTCCTTTGGCTAGGGAAGTGTT	caacAACACTTCCCTAGCCAAAGGAGC
gRNA3.2_Nme2Cas9_USH2A	caccGTCTCCAGAATCACACAAGTTAA	caacTTAACTTGTGTGATTCTGGAGAC
gRNA3.3_Nme2Cas9_USH2A	caccGATATGCAGGAGAAATAAAAGCC	caacGGCTTTTATTTCTCCTGCATATC
gRNA2_Nme2Cas9_HEKsite1 (#)	caccGAGTCAATGCAGATAGAGCTCTT	caacAAGAGCTCTATCTGCATTGACTC
gRNA3.1_Nme2Cas9_HEKsite1	caccGCTCCTCCCCTGAAACTACACGG	caacCCGTGTAGTTTCAGGGGAGGAGC
gRNA3.2_Nme2Cas9_HEKsite1	caccGAACTAAACCCCTATAACTTCA	caacTGAAGTTATAGGGGTTTAGTTC
gRNA3.3_Nme2Cas9_HEKsite1	caccGACTCCTGACCTCAGGTGATCCG	caacCGGATCACCTGAGGTCAGGAGTC
gRNA2_Nme2Cas9_BCR	caccGGCACCAGTTCTTGCCGTGCCCC	caacGGGGCACGGCAAGAACTGGTGCC
gRNA3.1_Nme2Cas9_BCR	caccGGGTTTCCTGTCATAACATAGG	caacCCTATGTTATGACAGGAAACCC
gRNA3.2_Nme2Cas9_BCR	caccGCTTACAAACACACACATCCCAC	caacGTGGGATGTGTGTGTTTGTAAGC
gRNA3.3_Nme2Cas9_BCR	caccGAACTCGTGTGTGAAACTCCAGA	caacTCTGGAGTTTCACACACGAGTTC
gRNA2_Nme2Cas9_TRBC	caccGGGCTGTCAGAGGAAGCTGGTC	caacGACCAGCTTCCTCTGACAGCCC
gRNA2_Nme2Cas9_TRAC	caccGGTGGCAATGGATAAGGCCGAG	caacCTCGGCCTTATCCATTGCCACC
gRNA3.1_Nme2Cas9_TRAC	caccGCCTCTTGTCCCACAGATATCCA	caacTGGATATCTGTGGGACAAGAGGC
gRNA3.2_Nme2Cas9_TRAC	caccGTTTGTCTGTGATATACACATCA	caacTGATGTGTATATCACAGACAAAC
gRNA3.3_Nme2Cas9_TRAC	caccGAACAGCCTGCGAAGGCACCAAA	caacTTTGGTGCCTTCGCAGGCTGTTC
gRNA2_Nme2Cas9_AAVS1	caccGCCCGAATCCACAGGAGAACGGG	caacCCCGTTCTCCTGTGGATTCGGGC
gRNA2_Nme2Cas9_FANCF	caccGCGGGCCAGGCTCTCTTGGAGT	caacACTCCAAGAGAGCCTGGCCCGC

gRNA3.1_Nme2Cas9_RHO	caccGCCATGGCAGTTCTCCATGCTG	caacCAGCATGGAGAACTGCCATGGC
gRNA3.2_Nme2Cas9_RHO	caccGAACTTCCTCACGCTCTACGTCA	caacTGACGTAGAGCGTGAGGAAGTTC
gRNA3.3_Nme2Cas9_RHO	caccGCTAGCCGTGGCTGACCTCTTCA	caacTGAAGAGGTCAGCCACGGCTAGC
BEgRNA1_Nme2Cas9_BCLenh	caccGCGGCATGGCATACAAATTATTT	caacAAATAATTTGTATGCCATGCCGC
BEgRNA2_Nme2Cas9_BCLenh	caccGTAACAGACACACGTATGTGTTG	caacCAACACATACGTGTGTCTGTTAC
BEgRNA3_Nme2Cas9_BCLenh	caccGCAGATAGCTGATTCCAGTGCA	caacTGCACTGGAATCAGCTATCTGC
BEgRNA1_Nme2Cas9_HEKsite1	caccGAGCACAAAGTGGGAAAGACCCA	caacTGGGTCTTTCCCACTTTGTGCTC
BEgRNA2_Nme2Cas9_HEKsite1	caccGAATCAATAGCTATGGTCAGATA	caacTATCTGACCATAGCTATTGATTC
BEgRNA1_Nme2Cas9_HEKsite2	caccGACACAAAGCATAGACTGCGGGG	caacCCCCGCAGTCTATGCTTTGTGTC
BEgRNA2_Nme2Cas9_HEKsite2	caccGACCAGTTTTATGAAGAAAAATG	caacCATTTTTCTTCATAAAACTGGTC
BEgRNA3_Nme2Cas9_HEKsite2	caccGTACTGACTACAGATATACCCT	caacAGGGTATATCTGTAGTCAGTAC
gRNA2_SaCas9_ATM	caccGGAACTTTATTGGCTGGAAC	aaacGTTCCAGCCAATAAAGTTCC
gRNA2_SaCas9_CHR6	caccGATTCTTACAACAACATGAGAG	aaacCTCTCATGTTGTTGTAAGAATC
gRNA2_SaCas9_NF1	caccGAGGGGTCTCCTTGTGGTGG	aaacCCACCACAAGGAGACCCCTC
gRNA2_SaCas9_DNMT1	caccGTGACACCGGGCGTGTTCCC	aaacGGGAACACGCCCGGTGTCAC
gRNA3.1_SaCas9_DNMT1	caccGCGGTGTCACGCCACTTGACAG	aaacCTGTCAAGTGGCGTGACACCGC
gRNA3.2_SaCas9_DNMT1	caccGACACTCCTCAAACGGTCCCCA	aaacTGGGGACCGTTTGAGGAGTGTC
gRNA3.3_SaCas9_DNMT1	caccGCGTTTCCCTCACTCCTGCTCG	aaacCGAGCAGGAGTGAGGGAAACGC
gRNA2_SaCas9_USH2A	caccGGTTTTACTATTCAACAATT	aaacAATTGTTGAATAGTAAAACC
gRNA3.1_SaCas9_USH2A	caccGCAGGCCAGTTGATTTGTATA	aaacTATACAAATCAACTGGCCTGC
gRNA3.2_SaCas9_USH2A	caccGATTCGCTGCTCTTGTTGCAG	aaacCTGCAACAAGAGCAGCGAATC
gRNA3.3_SaCas9_USH2A	caccGCAGTTTGTTCTCTGTGATTG	aaacCAATCACAGAGAACAAACTGC
gRNA2_SaCas9_HEKsite1	caccGCTCTTGGTACCTGAAGTTAT	aaacATAACTTCAGGTACCAAGAGC
gRNA3.1_SaCas9_HEKsite1	caccGCTTTTCAGCGACCCACGGATG	aaacCATCCGTGGGTCGCTGAAAAGC
gRNA3.2_SaCas9_HEKsite1	caccGGCAATAGAGGCAAGTATGAG	aaacCTCATACTTGCCTCTATTGCC
gRNA3.3_SaCas9_HEKsite1 (#)	caccGCACTTTGAGAGGCCGAGGTGG	aaacCCACCTCGGCCTCTCAAAGTGC
gRNA2_SaCas9_BCR	caccGTTCTTGCCGTGCCCCTTCCC	aaacGGGAAGGGGCACGGCAAGAAC
gRNA3.1_SaCas9_BCR	caccGAGAGGTGGCTCTGCATAGGT	aaacACCTATGCAGAGCCACCTCTC
gRNA3.2_SaCas9_BCR	caccGAGTGCAGTGCTGGTCTGGCGG	aaacCCGCCAGACCAGCACTGCACTC
gRNA3.3_SaCas9_BCR	caccGGACAGTCTGGAGTTTCACAC	aaacGTGTGAAACTCCAGACTGTCC
gRNA2_SaCas9_TRBC	caccGAGGAAGCTGGTCTGGGCCT	aaacAGGCCCAGACCAGCTTCCTC

gRNA2_SaCas9_TRAC	caccGAGGCCGAGACCACCAATCAG	aaacCTGATTGGTGGTCTCGGCCTC
gRNA3.1_SaCas9_TRAC	caccGAGTCTCTCAGCTGGTACACGG	aaacCCGTGTACCAGCTGAGAGACTC
gRNA3.2_SaCas9_TRAC	caccGTTTTGTCTGTGATATACACAT	aaacATGTGTATATCACAGACAAAAC
gRNA3.3_SaCas9_TRAC	caccGCGCAGGCTGTTTCCTTGCTTC	aaacGAAGCAAGGAAACAGCCTGCGC
gRNA2_SaCas9_AAVS1	caccGACCCGAATCCACAGGAGAA	aaacTTCTCCTGTGGATTCGGGTC
gRNA2_SaCas9_FANCF	caccGCAAGGCGGGCCAGGCTCTCT	aaacAGAGAGCCTGGCCCGCCTTGC
gRNA2_SaCas9_VEGFAsite3	caccGATTCCCTCTTTAGCCAGAGC	aaacGCTCTGGCTAAAGAGGGAATC
gRNA2_SaCas9_PD1	caccGCTCTGGGCCCTGCTGGGTAC	aaacGTACCCAGCAGGGCCCAGAGC
gRNA3.1_SaCas9_RHO	caccGTGGCTCAGCCAGGTAGTACT	aaacAGTACTACCTGGCTGAGCCAC
gRNA3.2_SaCas9_RHO	caccGTCAGCCACGGCTAGGTTGAG	aaacCTCAACCTAGCCGTGGCTGAC
gRNA3.3_SaCas9_RHO	caccGCACCCTCTACACCTCTCTGC	aaacGCAGAGAGGTGTAGAGGGTGC
BEgRNA1_SaCas9_BCLenh (BCL-BEg1)	caccGAATTGGATTTTATTTCTCAAT	aaacATTGAGAAATAAAATCCAATTC
BEgRNA2_SaCas9_BCLenh (BCL-BEg2)	caccGAGCTCCAAACTCTCAAACCAC	aaacGTGGTTTGAGAGTTTGGAGCTC
BEgRNA3_SaCas9_BCLenh (BCL-BEg3)	caccGTTACCTGTATGGACTTTGCAC	aaacGTGCAAAGTCCATACAGGTAAC
BEgRNA4_SaCas9_BCLenh (BCL-BEg4)	caccGATAATAATAGTATATGCTTCA	aaacTGAAGCATATACTATTATTATC
BEgRNA1_SaCas9_HEKsite2 (H2-BEg1)	caccGCTGAATAGCTGCAAACAAGTG	aaacCACTTGTTTGCAGCTATTCAGC
BEgRNA2_SaCas9_HEKsite2 (H2-BEg2)	caccGAATGATATCCATTATTAGTGG	aaacCCACTAATAATGGATATCATTC
BEgRNA3_SaCas9_HEKsite2 (H2-BEg3)	caccGTCTCCCAAAACATGCCCATAT	aaacATATGGGCATGTTTTGGGAGAC
BEgRNA4_SaCas9_HEKsite2 (H2-BEg4)	caccGCTGAGGGGAGCAGTGGGGGA	aaacTCCCCCACTGCTCCCCTCAGC
BEgRNA1_SaCas9_HEKsite3 (H3-BEg1)	caccGTGGAGCTTGCCCAGGACCCCA	aaacTGGGGTCCTGGGCAAGCTCCAC
BEgRNA1_SaCas9_HEKsite4 (H4-BEg1)	caccGAGGGGAGGGGAAAGAAGACCA	aaacTGGTCTTCTTTCCCCTCCCCTC
BEgRNA1_SaCas9_TRAC (TRAC-BEg1)	caccGCTCACTGTTTCTTAGTAAAAA	aaacTTTTTACTAAGAAACAGTGAGC
gRNA3.1_CjCas9_BCR	accgCCTATGTTATGACAGGAAACCC	gaacGGGTTTCCTGTCATAACATAGG
gRNA3.2_CjCas9_BCR	accgAGTCTTAGCATGAGACTTACAA	gaacTTGTAAGTCTCATGCTAAGACT
gRNA3.3_CjCas9_BCR	accgGCCCCTCCCACTGGACCTCCA	gaacTGGAGGTCCAGTGGGAGGGGC
gRNA3.1_CjCas9_DNMT1	accgAGCCAAGGCCACAAACACCATG	gaacCATGGTGTTTGTGGCCTTGGCT

gRNA3.2_CjCas9_DNMT1	accgCTGTTACTCGCCTGTCAAGTGG	gaacCCACTTGACAGGCGAGTAACAG
gRNA3.3_CjCas9_DNMT1	accGACATGGACCATCAGGAAACAT	gaacATGTTTCCTGATGGTCCATGT
gRNA3.1_CjCas9_HEKsite1	accgGGGGCGTGCCGGCGCATGGCA	gaacTGCCATGCGCCGGCACGCCCC
gRNA3.2_CjCas9_HEKsite1	accgAAGACGCTGGTCTAGAGACCCA	gaacTGGGTCTCTAGACCAGCGTCTT
gRNA3.3_CjCas9_HEKsite1	accgGCGTGAGCCACCAGGCCCGGG	gaacCCCGGGCCTGGTGGCTCACGC
gRNA3.1_CjCas9_TRAC	accgCTTGAAGTCCATAGACCTCATG	gaacCATGAGGTCTATGGACTTCAAG
gRNA3.2_CjCas9_TRAC	accgAGTCAGATTTGTTGCTCCAGGC	gaacGCCTGGAGCAACAAATCTGACT
gRNA3.3_CjCas9_TRAC	accgCTTCAACAACAGCATTATTCCA	gaacTGGAATAATGCTGTTGTTGAAG
gRNA3.1_CjCas9_RHO	accGTACCACACCCGTCGCATTGGAG	gaacCTCCAATGCGACGGGTGTGGTA
gRNA3.2_CjCas9_RHO	accGCGTGAGGAAGTTGATGGGGAAG	gaacCTTCCCCATCAACTTCCTCACG
gRNA3.3_CjCas9_RHO	accgCTTCATGGTCCTAGGTGGCTTC	gaacGAAGCCACCTAGGACCATGAAG
gRNA3.1_CjCas9_USH2A	accgTAGATGATTCGGCTTATCATTT	gaacAAATGATAAGCCGAATCATCTA
gRNA3.2_CjCas9_USH2A	accGAAACAAATCATGAAATTGAAA	gaacTTTCAATTTCATGATTTGTTT
gRNA3.3_CjCas9_USH2A	accgTCATTATCATGGATGTTTCACC	gaacGGTGAAACATCCATGATAATGA
BEgRNA1_CjCas9_BCLenh	accgAATTTTGGGAGTCCACACGGCA	gaacTGCCGTGTGGACTCCCAAAATT
BEgRNA2_CjCas9_BCLenh	accgCAAACTCTCAAACCACAGGGAT	gaacATCCCTGTGGTTTGAGAGTTTG
BEgRNA3_CjCas9_BCLenh	accgAGATAGCTGATTCCAGTGCAAA	gaacTTTGCACTGGAATCAGCTATCT
BEgRNA1_CjCas9_HEKsite1	accgGAGAGCCGTGTAGTTTCAGGGG	gaacCCCCTGAAACTACACGGCTCTC
BEgRNA2_CjCas9_HEKsite1	accgATGTGCAAATCAATAGCTATGG	gaacCCATAGCTATTGATTTGCACAT
BEgRNA3_CjCas9_HEKsite1	accgCTGAGTCAATGCAGATAGAGCT	gaacAGCTCTATCTGCATTGACTCAG
BEgRNA1_CjCas9_HEKsite2	accgCCGCTGGCCCTGTAAAGGAAAC	gaacGTTTCCTTTACAGGGCCAGCGG
BEgRNA2_CjCas9_HEKsite2	accgTATCATTTTTCTTCATAAAACT	gaacAGTTTTATGAAGAAAAATGATA
BEgRNA3_CjCas9_HEKsite2	accgACAGATATACCCTAACACCAAC	gaacGTTGGTGTTAGGGTATATCTGT
BEgRNA1_SauriCas9_BCLenh	caccGATAATTTGTATGCCATGCCGT	aaacACGGCATGGCATACAAATTATC
BEgRNA3_SauriCas9_BCLenh	caccGAAGTCCATACAGGTAATAACA	aaacTGTTATTACCTGTATGGACTTC
BEgRNA1_SauriCas9_HEKsite1	caccGTGGGAAAGACCCAGCATCCG	aaacCGGATGCTGGGTCTTTCCCAC
BEgRNA2_SauriCas9_HEKsite1	caccGCCCTGAACATGTGGTACACAA	aaacTTGTGTACCACATGTTCAGGGC
BEgRNA1_SauriCas9_HEKsite2	caccGACACAAAGCATAGACTGCGGG	aaacCCCGCAGTCTATGCTTTGTGTC
BEgRNA2_SauriCas9_HEKsite2	caccGAAAAATGATATCCATTATTA	aaacTAATAATGGATATCATTTTTC
gRNA3.1_AsCas12f1_BCR	gaacGGGGCCGAGAGGTGGCTCT	aaaaAGAGCCACCTCTCGGCCCC
gRNA3.2_AsCas12f1_BCR	gaacGGTAGCGTGGGATGTGTGT	aaaaACACACATCCCACGCTACC

gRNA3.1_AsCas12f1_DNMT1	gaacGCAGCTTCCTCCTCCTTTA	aaaaTAAAGGAGGAGGAAGCTGC
gRNA3.2_AsCas12f1_DNMT1	gaacACGTACTGATGTTAACAGC	aaaaGCTGTTAACATCAGTACGT

(*) As for the guides used for the comparison of CoCas9 with SpCas9, Nme2Cas9 and SaCas9, gRNA1 indicates guides NOT overlapping with SpCas9/SaCas9/Nme2Cas9 guides, while gRNA2 indicates overlapping guides. If gRNA2 is not indicated, gRNA1 is overlapping with SpCas9/SaCas9/Nme2Cas9. gRNA3 indicates NOT OVERLAPPING guides used in the comparison between CoCas9/CjCas9/Nme2Cas9/SaCas9.

(**) BEgRNA indicates guides employed for base editing experiments.

(#) This guide was also used in base editing experiments.

(##) The cloning overhang is reported in lowercase.

Table 6. Sequences of the sgRNAs of the Cas9 orthologs presented in this work.

CoCas9 Full scaffold (FS)	NNNNNNNNNNNNNNNNNNNNGUCUUGAGCACGCGCCCUUCCCCAAGGUGAGAAAUC ACCUUGGGGAAGGGCGCGGCUCCAGACAAGGGGAGCCACUUAAGUGGCUUACCCG UAAAGUAACCCCCGUUCAAUCUUCGGAUUGGGCGGGGCGAACUUUUUU
CoCas9 Full scaffold optimized (FS-opt)	NNNNNNNNNNNNNNNNNNNNGUCUUGAGCACGCGCCCUUCCGCAAGGUGAGAAAUC ACCUUGCGGAAGGGCGCGGCUCCAGACAAGCGGAGCCACUUAAGUGGCUUACGCG UAAAGUAACCGCCGUUCAAUCUUCGGAUUGGGCGGCGCGAACUUUUUU
CoCas9 trimmed scaffold (TS)	NNNNNNNNNNNNNNNNNNNNGUCUUGAGCACGCGAAAGCGGCUCCAGACAAGGGGA GCCACUUAAGUGGCUUACCCGUAAAGUAACCCCCGUUCAAUCUUCGGAUUGGGCGG GGCGAACUUUUUU
CoCas9 trimmed scaffold optimized (TS-opt)	NNNNNNNNNNNNNNNNNNNNGUCUUGAGCACGCGAAAGCGGCUCCAGACAAGCGGA GCCACUUAAGUGGCUUACGCGUAAAGUAACCGCCGUUCAAUCUUCGGAUUGGGCG GCGCGAACUUUUUU
CaCas9 Full scaffold	NNNNNNNNNNNNNNNNNNNNGUCUUGAGCACGCGCCCUUCCCCAAGGUGAGAAAUC ACCUUGGGGAAGGGCGCGGCUCCAGACAAGGGGAGCCACUUAAGUGGCUUACCCG UAAAGUAACCCCCGUUCAAUCUUCGGAUUGGGCGGGGCGAACUUUUUU
Al1Cas9 Full scaffold	NNNNNNNNNNNNNNNNNNNNGUUAUAGCUUCCUUUCCAAAUCAGACAUGCUAUAGA AAUAUUAUAUGUCUGAUUUGGAAAGGAAGUCUAUAAUAAUCGAAGUUAUCUUUACGA GUAGGGCUCUGACGUCACAUAUAAUAUAUGUGGCGUCAUCCUUUUUU
GeCas9 Full scaffold	NNNNNNNNNNNNNNNNNNNNGUCAUAGUUCCCUAAUAGCUCUUGGUAUGGUAUAAU GAAAUUAUACCAUACCAAGAACUAUUAUGGUUGCUAUGAUAAGGUCAUAGGACCGUA AAGCUCUGACGCCCUGUCCUAGGACAGGGCGUCAUCUUUUUUUUUUUU
Al3Cas9 Full scaffold	NNNNNNNNNNNNNNNNNNNNGUUAUAGGUUGCUUCUCAAAUCGUCGAAAGACGAUU UGAGAAGCAACCGUUUAUAUAAUAAAGAAUAUAUUCUGUGGGGUUCUACGAUUAUG UAUCCCCUGAAACCGUCCCCCUGCGGGGACGGUUUCUGUUUUUU
ClCas9 Full scaffold	NNNNNNNNNNNNNNNNNNNNAUUGUAGUUCCCUGAAAGGGUUACUACAAAAAGGUA GAAAACCGAAAAGCUCUGACGGCUCCUUAUUGGAGCCGUUAUCUUUUUUUUUU
SuCas9 Full scaffold	NNNNNNNNNNNNNNNNNNNNGUUCUGGCUUUAGCUCGUUUCCUAACGAUAUAGAAA UAUAUCGUUAGGAAACGAGCUAUUGCCACUAACAAGCAUUGCAAAAUGCAACAGAUC CGAGGGGUGCUAUAUGCGCCCCUUAUUUCUUUCUGAAUUUUUU
BsCas9 Full scaffold	NNNNNNNNNNNNNNNNNNNNAUCAUAGUUCCCUAAGAAAUUAUGGUUGCUAUGAUA AGGUCAUAGGACCGUAAAGCUCUGACGCCCUGCUAUUUGGCAGGGCGUCAUCUUU UUU

Al4Cas9 Full scaffold	NNNNNNNNNNNNNNNNNNNNGUUAUAGGUUGCUUCUCAAAUCGUCGAAAGACGAUU UGAGAAGCAACCGUUUAUAUAAUAAAGAAUAUAUUCUGUGGGGUUCUACGAUUAUG UAUCCCCUGAAGCCGUCCCCCUGCGGGGACGGUUUCUGUUGUAUUAUUUUUU
Al2Cas9 Full scaffold	NNNNNNNNNNNNNNNNNNNNGUUAUAGCUUCCUUUCCAAAUCAGAUAUGCUAUGAA AAUUAUAUAUCUGAUUUGGAAAGGAAGCUAUAAUAAUCGAAAUUCGAGUGGGGUAC UGAUGAUAGUCUAAAGCAAUUUAGAUGUGUAGUCAUCCCCUUUUUU
TmCas9 Full scaffold	NNNNNNNNNNNNNNNNNNNNNNNNGTCTTGAGCACGCACCCTTCCCCAAGGTGAGAAA TCACCTTGGGGAAGGGTGCGGCTCCAGACAAGGGAAGTCAGCTATCTGACTTACCCGT AAAGTTACCCCCGCACCGTCCTCGGACGATGCGGGGCGAACTTTTTT