

RESEARCH

Open Access

Eukaryotic-driven directed evolution of Cas9 nucleases



Giulia Vittoria Ruta^{1*}, Matteo Ciciani^{1,2}, Eyemen Kheir¹, Michele Domenico Gentile¹, Simone Amistadi^{1,4}, Antonio Casini³ and Anna Cereseto^{1*} 

*Correspondence: giuliovittoria.ruta@unitn.it; anna.cereseto@unitn.it

¹ Laboratory of Molecular Virology, Department CIBIO, University of Trento, Trento, Italy

² Laboratory of Computational Metagenomics, Department CIBIO, University of Trento, Trento, Italy

³ Alia Therapeutics S.R.L, Trento, Italy

⁴ Present address: Laboratory of Chromatin and Gene Regulation During Development, Université de Paris, Imagine Institute, INSERM UMR 1163, Paris, France

Abstract

Background: Further advancement of genome editing highly depends on the development of tools with higher compatibility with eukaryotes. A multitude of described Cas9s have great potential but require optimization for genome editing purposes. Among these, the Cas9 from *Campylobacter jejuni*, CjCas9, has a favorable small size, facilitating delivery in mammalian cells. Nonetheless, its full exploitation is limited by its poor editing activity.

Results: Here, we develop a Eukaryotic Platform to Improve Cas Activity (EPICA) to steer weakly active Cas9 nucleases into highly active enzymes by directed evolution. The EPICA platform is obtained by coupling Cas nuclease activity with yeast auxotrophic selection followed by mammalian cell selection through a sensitive reporter system. EPICA is validated with CjCas9, generating an enhanced variant, UltraCjCas9, following directed evolution rounds. UltraCjCas9 is up to 12-fold more active in mammalian endogenous genomic loci, while preserving high genome-wide specificity.

Conclusions: We report a eukaryotic pipeline allowing enhancement of Cas9 systems, setting the ground to unlock the multitude of RNA-guided nucleases existing in nature.

Background

The development of CRISPR-Cas tools [1, 2] and the emerging number of RNA-guided prokaryotic and eukaryotic nucleases [3–5] are giving a tremendous impulse to the development of advanced therapies based on genomic modifications [6]. Yet, to further accelerate and broaden the application of genome editing [7], a larger set of tools is highly needed to match the complexity of gene therapy approaches. In particular, there is a demand to increase the accessibility to any genomic site by enlarging the repertoire of enzymes covering the largest number of protospacer adjacent motifs (PAM) [8, 9]. Moreover, the efficacy of cellular and organ delivery, which highly depends on the cargo size [10], is very relevant to the field, thus reinforcing the need for low molecular weight Cas enzymes. Small molecular size is indeed necessary for Cas compatibility with an adeno-associated viral vector (AAV) without enzyme splitting [11] and allows



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

the generation of compact base editors and prime editors, which can precisely edit DNA avoiding double-strand breaks [1, 2].

A large reservoir of CRISPR-Cas systems and novel emerging ancestral RNA-guided nucleases have been identified in metagenomic data banks. Nonetheless, the majority of enzymes isolated from bacteria [12], archaea [13], and viruses [14] are poorly active in mammalian cells. Strategies aimed at enhancing these systems for genome editing include molecular engineering of the nucleases by rational design [15], bacterial screening [16], or optimization of the sgRNA [17]. Despite these efforts, few CRISPR-Cas tools are sufficiently active for efficient genome editing in eukaryotes [18], as demonstrated by the paucity of RNA-guided nucleases for experimental use, as opposed to the multitude of systems identified in metagenomic data banks. To unlock the natural reservoir of prokaryotic systems, experimental setups are needed to enable the adaptation of enzymes to the eukaryotic nuclear environment. We have previously demonstrated that yeast cells are ductile eukaryotic systems that can improve Cas precision by directed evolution [19]. On this groundwork, we generated a Eukaryotic Platform to Improve Cas Activity (EPICA). To demonstrate the efficacy of this evolution platform, we introduced EPICA CjCas9 [20], a small size (984 aa) Cas enzyme with favorable features for genome editing, which is however minimally used due to its reduced nuclease activity [21]. The platform produced an enhanced variant, UltraCjCas9, carrying modifications hardly predictable through a rational engineering approach.

Results

Yeast-based directed evolution enhances Cas9 activity

To generate a eukaryotic platform for the evolution of Cas9 nuclease activity, a yeast strain was engineered by integrating two different cleavage reporter cassettes in the tryptophan (TRP) and adenine (ADE) genes. Disruption of the TRP and ADE genes by cassette insertion turns yeast growth dependent from adenine and tryptophan culture integration. The cassettes contain target sites including different PAM sequences compatible with the Cas9 under evolution, to allow the isolation of variants evolved independently from specific PAM sequence recognition. Homology arms were inserted at the edges of the cassettes to mediate single-strand annealing (SSA) repair induced by Cas9 cleavage of the target locus leading to gene reconstitution [19]. This model, schematized in Fig. 1a, allows the selection of active Cas9 molecules by auxotrophy through limiting growth factors conditions (TRP and ADE deprivation, TRP – and ADE –, respectively), permitting exclusive proliferation of cells with restored TRP and ADE loci generated by efficient Cas9 cleavage. The platform was initially tested with SpCas9, showing high levels of yeast cell survival in single ADE – and TRP – or combined ADE – /TRP – conditions; by contrast, wild-type CjCas9 (CjCas9 WT) showed 20% of living colonies in ADE – conditions but nearly absent survival in ADE/TRP deprived culture conditions (Fig. 1b). Transformation of a cDNA library of CjCas9 variants obtained through random PCR mutagenesis led to isolation of colonies growing in ADE – /TRP – conditions (Fig. 1c). Starting from the colonies obtained by transformation of the mutagenized CjCas9 library, the process of random mutagenesis and transformation in ADE – /TRP – conditions was reiterated until a decreasing trend in cell survival was observed (rounds 5 and 6), likely resulting from saturating mutagenesis. The cDNA isolated from

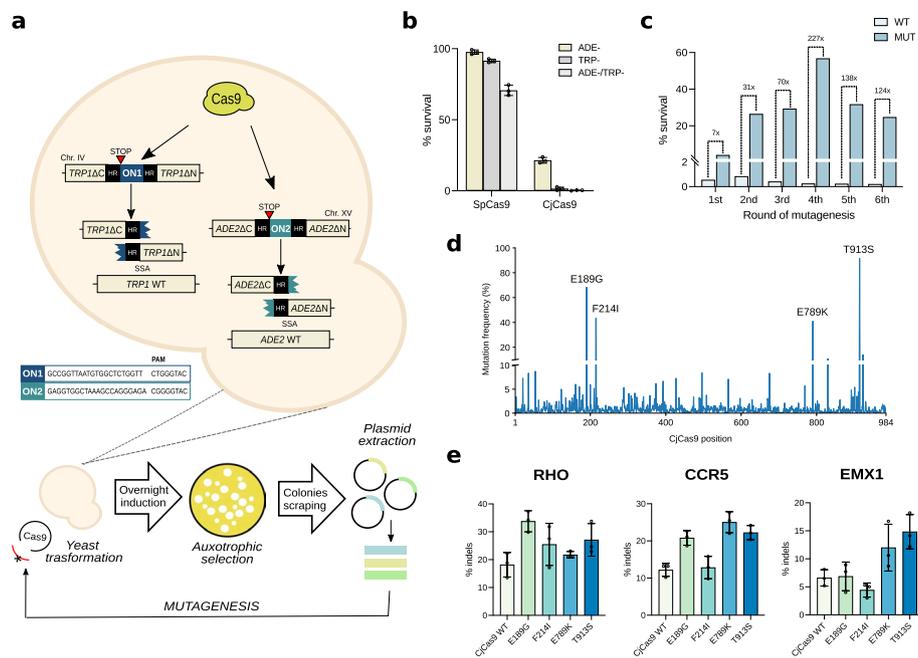


Fig. 1 Directed evolution step to enhance Cas nuclease activity. **a** Schematic and experimental flowchart of the yeast Cas evolution step. Framed sequences correspond to the ON1 and ON2 cassettes carrying the CjCas9 target sites and AAVS1-TS34 and AAVS1-TS32 targets, respectively. Active mutants cleaving both ON1 and ON2 in the ADE2 and TRP1 loci are amplified from colonies growing in ADE – and TRP – conditions and undergo additional rounds of mutagenesis and selection. **b** Cleavage activity of SpCas9 and CjCas9 are reported as percentages of survival, corresponding to the number of colonies growing in TRP – and ADE – plates with respect to total transformants measured in unrestricted growth conditions. Culture plates shown in Additional file 1: Fig. S1a. **c** The activity of the CjCas9 variants (MUT) isolated at each round of evolution was compared with CjCas9 WT by measuring the percentages of survival after transformation. Culture plates are shown in Additional file 1: Fig. S1b. **d** Peaks at each CjCas9 amino acid position represent the frequency of the mutated residues obtained by NGS deep sequencing. Amino acid substitutions with high frequency are reported in the graph. **e** Editing activity at the endogenous loci in HEK293 cells of CjCas9 variants containing substitutions corresponding to the most frequently mutated amino acids in **d**. In **b** and **e**, data are reported as mean \pm standard deviation of $n \geq 3$ biologically independent samples. Individual values are represented as empty circles

the fourth round of mutagenesis, where the highest degree of cleavage activity was detected (227-fold compared to wild-type, Fig. 1c), was deep sequenced revealing a large variety of substitutions throughout the nuclease sequence (Fig. 1d). As an intermediate step to verify the yeast evolution step, we sampled the cDNA library to evaluate genome editing improvement in mammalian cells. We selected the most frequent substitutions (E189G, F214I, E789K, T913S, Fig. 1d) which were tested individually for indel formation in three endogenous loci (*RHO*, *CCR5*, and *EMX1*) of HEK293 cells. Each single mutation showed variable editing improvement (Fig. 1e), thus suggesting that enhancement in yeast corresponded to optimization in a mammalian context.

Selection of the enhanced CjCas9 variants in a mammalian reporter cell line

To select the best combinations of mutations for CjCas9 enhancement in mammalian cells, we generated a reporter cell line carrying a switch-on EGFP circuit controlled by indel formation. HEK293 cells carrying a Tet repressor (TetR) cDNA were engineered with a EGFP cassette repressed by Tet operator (TetO) elements, which

are released by CjCas9 cleavage of TetR (Fig. 2a and Additional file 1: Fig. S2a). The Cas-mediated induction of fluorescence increases the dynamic range compared to the widely used EGFP disruption assay [19], thus facilitating the detection of differences in Cas activity (Additional file 1: Fig. S2a). A lentiviral library of the yeast-evolved CjCas9 variants was transduced in the TetR-EGFP reporter cell line to select the most active mutants (Additional file 1: Fig. S2b). The selection step was repeated a second time to further enrich the best variants, which showed almost tenfold higher fluorescence than controls, indicating the presence of mutants more active than the original enzyme (Fig. 2b and Additional file 1: Fig. S2b). A large number of variants ($n = 160$) were isolated from double-sorted EGFP cells and tested individually (Additional file 1: Fig. S2c). We selected the best-performing variants, corresponding to mutants having the highest ratio of positive EGFP cells with respect to CjCas9 WT in the TetR-EGFP reporter cell line. The best-performing variants ($n = 17$) were further evaluated for indel formation in the *EMX1*, *RHO*, and *CCR5* endogenous loci (Fig. 2c), revealing an average increased activity compared to wild-type ranging from 2- up to 3.6-fold (Additional file 2: Table S1). Interestingly, the activity fold change observed at the *EMX1* locus (mean 4.4-fold) is consistently larger compared to *RHO* and *CCR5* (mean 1.9-fold for both), possibly because the former locus is actively expressed in HEK293 cells, while the latter loci are not [22].

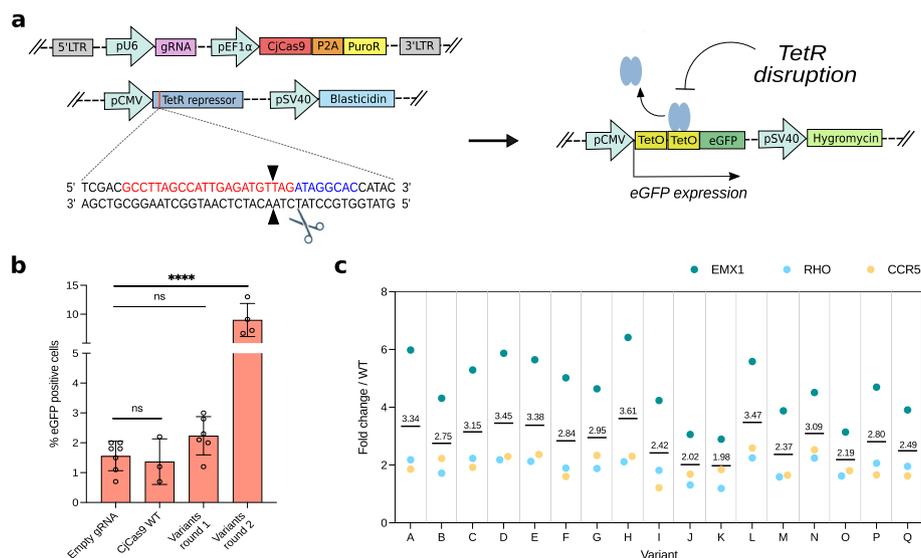


Fig. 2 CjCas9 enhancement through screening in mammalian cells. **a** Schematic representation of the TetR-EGFP reporter cell line. TetR (blue) is constitutively expressed inactivating EGFP (green) expression. TetR inactivation by Cas9 cleavage enables EGFP expression and isolation of fluorescent cells by sorting. **b** CjCas9 variants isolated from rounds 1 and 2 of the mammalian screening are compared with CjCas9 WT transducing the TetR-EGFP reporter cell line. Statistical analysis was performed using one-way ANOVA with Tukey's multiple comparisons test: ns, not significant, **** $P < 0.0001$. **c** Editing activity of selected variants obtained from yeast and mammalian selection. Dots represent the fold changes of activity of each variant (from A to Q) with respect to CjCas9 WT targeting EMX, RHO, or CCR5. Extended data are shown in Additional file 1: Fig. S3. The grand mean values of each variant are shown in the graph. Sanger sequences of the variants are reported in Additional file 2: Table S2. In **b**, data are reported as mean \pm standard deviation of $n \geq 3$ biologically independent samples. Individual values are represented as empty circles

In conclusion, through the evolution pressure obtained by coupling Cas cleavage activity with auxotrophic selection in yeast and indel formation in mammalian cells, EPICA enables the generation of enhanced Cas variants.

UltraCjCas9, an evolved highly active Cas9

The most active variant, named UltraCjCas9 hereafter, was further characterized. Sanger sequencing revealed five modifications in four domains (RuvC-I, REC1, HNH, and PI) (Fig. 3a and Additional file 1: Fig. S4). Given the presence of one mutation in the PI domain (T913S), we verified the PAM preference of UltraCjCas9 through an in vitro PAM determination assay. The results showed that UltraCjCas9 recognizes a N₄RYAY consensus sequence, relaxing positions 4 and 8 of the wild-type N₃VRYAC, thus expanding the targeting range (Fig. 3b and Additional file 1: Fig. S5). To evaluate the editing profile, we performed a comparative editing analysis in 23 human genomic loci with UltraCjCas9, CjCas9 WT, and the recently optimized enCjCas9 [15] showing that UltraCjCas9 was more active in the majority of sites (Fig. 3c). The editing efficacy of UltraCjCas9 was quite variable, reaching up to 12-fold higher activity than the wild-type (*HEK site 1*) (Fig. 3c and Additional file 2: Table S3). Overall, UltraCjCas9 was more active than CjCas9 WT (28.1% and 13.0% mean editing, respectively) and outperformed enCas9 (17.5% mean editing), which was obtained by rational engineering (Fig. 3d). Given the editing efficacy of UltraCjCas9, we compared this variant with the most commonly used Cas9 from *Streptococcus pyogenes*, SpCas9, by selecting genomic loci efficiently edited by UltraCjCas9 and using overlapping guides for both orthologs (Additional file 1: Fig. S6b). Overall, the two nucleases showed no significant difference in editing efficiency (Fig. 3e).

Finally, since enhanced Cas9 nuclease activity and relaxed PAM may correlate with decreased fidelity [23, 24], we performed a whole-genome off-target analysis by GUIDE-Seq [25] to compare the specificity of CjCas9 WT and UltraCjCas9. We chose two sites, *AAVSI-TS1* and *AAVSI-TS8*, previously associated with CjCas9 off-target activity [20], where both nucleases produced similar levels of on-target indels (Additional file 1: Fig. S6c). In both genomic loci, UltraCjCas9 generated less off-target cleavage sites than the wild-type enzyme, resulting in slightly higher precision (Fig. 3f). The similar precision profile was confirmed by comparing the sgRNA mismatch tolerance (single nucleotide mismatches) of CjCas9 WT and UltraCjCas9 (Additional file 1: Fig. S7). Overall, UltraCjCas9 exhibited a specificity profile similar to CjCas9 WT, confirming that the enhanced activity did not alter editing precision.

Overall, UltraCjCas9, generated by EPICA, showed an enhanced activity and an expanded targeting range while preserving specificity.

Discussion

In just over a decade, CRISPR-Cas systems working as adaptive immunity in prokaryotes have been successfully translated into genome editing tools for mammalian cells up to the clinic. Notably, the prokaryotic cellular environment highly differs from the eukaryotic nuclear compartment, characterized by multiple factors defining chromatin structure and function. Hence, to further advance this technology, CRISPR-Cas systems should be functionally steered in eukaryotic cells. Along this line, emerging data from

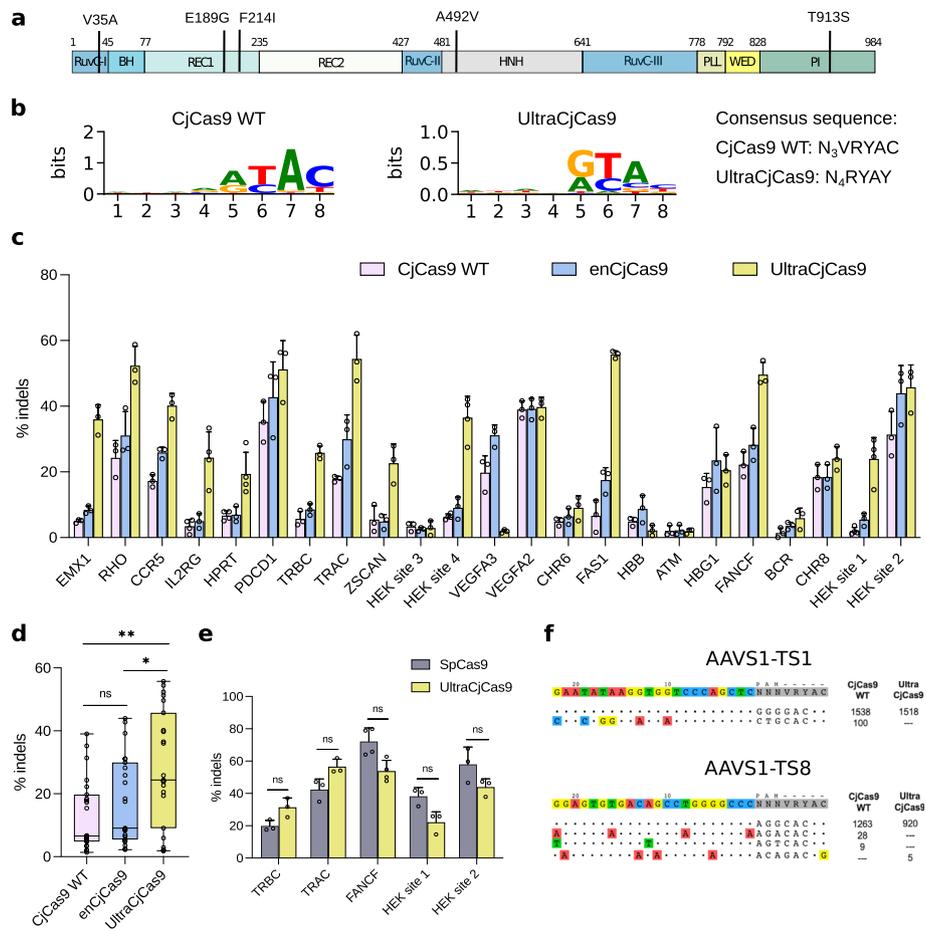


Fig. 3 Characterization of the highly enhanced variant UltraCjCas9. **a** Domain localization of the five amino acid substitutions (V35A, E189G, F214I, A492V, and T913S) characterizing UltraCjCas9. See Additional file 1: Fig. S4 for amino acid positions in the CjCas9 structure. **b** Sequence logo representations of the PAM preference of CjCas9 and UltraCjCas9, obtained using an in vitro PAM determination assay. Heatmaps representing each nucleotide combination are reported in Additional file 1: Fig. S5. **c** Editing efficiency in the endogenous genomic loci of HEK293 cells with CjCas9 WT, enCjCas9, and UltraCjCas9. PAM sequences are reported in Additional file 1: Fig. S6a. **d** Graphical summary of editing activities of CjCas9 WT (pink), enCjCas9 (blue), and UltraCjCas9 (yellow) in **c**. Empty circles in the box plots represent the average percentages of indels for each genomic locus. Statistical analysis was performed using one-way ANOVA with Tukey's multiple comparisons test: ns, not significant, * $P < 0.05$, ** $P < 0.01$. Central line, median; box limits, upper and lower quartiles; whiskers, $\times 1.5$ interquartile range; $n = 23$ independent loci. **e** Comparison of SpCas9 and UltraCjCas9 editing activity in the indicated endogenous genomic loci. Statistical significance was assessed using a two-sided t -test corrected using the Holm-Šidák method for each locus. ns, not significant. **f** GUIDE-Seq analysis performed with CjCas9 WT and UltraCjCas9. GUIDE-seq read counts of each on- and off-target in HEK293T cells are shown on the right side (extended data in Additional file 2: Table S7–S10). In **c** and **e**, data are reported as mean \pm standard deviation of $n \geq 3$ biologically independent samples. Individual values are represented as empty circles

various groups demonstrated that chromatin structure impacts CRISPR-Cas genome editing efficacy [26–34]. Remarkably, while nucleosomes affect the editing activity of CRISPR-Cas enzymes, the artificial zinc finger nucleases (ZFN) remain unaffected, suggesting a natural evolution of Cas nucleases to work with naked DNA [29]. More recently, direct evidence demonstrated the role of chromatin structure on Cas12a PAM recognition and R-loop formation efficiency [32] and the effect of chromatin remodeling

factors on the nuclease activity of a variety of Cas9 proteins, including small orthologs [33]. While the effect of chromatin structure on Cas efficiency has been well documented, other factors could influence nuclease activity in eukaryotic cells, such as interactions with host proteins. Post-translational modifications (PTMs) have been shown to inhibit Cas proteins in prokaryotes [34–36]; however, it remains unclear to what extent nuclease activity could be modulated by PTMs introduced by host factors in eukaryotes [37]. These results suggest that limited insights into Cas interactions with the eukaryotic environment will severely limit the molecular engineering of these tools.

While rational design has been widely employed to engineer Cas nucleases [15, 38–40], this approach requires protein structure information which is not necessarily available for the large number of RNA-guided nucleases requiring optimization as genome editing tools. Moreover, resolved protein structures are often missing flexible regions, such as the HNH domain of Cas9 nucleases [41–43], imposing limitations on the rational design approach. In general, enhancement of Cas9 nucleases by rational engineering is obtained through modifications of residues directly contacting the DNA-RNA hybrid, often increasing positive charges in order to facilitate interactions between the Cas9 protein and the negatively charged nucleic acid backbone [15, 39]. In line with this type of enhancing modifications, UltraCjCas9 carries the substitution E189G in the REC1 domain, which neutralizes a negative charge in the vicinity of the PAM-proximal region of the DNA-RNA hybrid. Conversely, the remaining mutations in UltraCjCas9 would have been hardly predictable through structural analysis, demonstrating the advantage of an unbiased in cellulo Cas9 selection over rational engineering [15]. In particular, mutations V35A and A492V are absent from the crystal structure of CjCas9 [41]. According to the AlphaFold prediction [44], V35A is located in a loop connecting the RuvC-I and BH domains, while A492V is located near the base of an alpha helix at the start of the HNH domain (Additional file 1: Fig. S4). Unexpectedly, these mutations are chemically similar to the wild-type residues, and their effect could be mediated by different steric hindrances of their side chains. Moreover, mutation F214I is located at the end of an alpha helix in the REC1 domain and is positioned far from DNA-RNA interactions in the crystal structure. Specifically, Phe214 is positioned near the cleft between REC1 and REC2, which accommodates the unmelted PAM-distal dsDNA during R-loop formation [45]. Finally, mutation T913S, located in the PAM interacting domain, is likely responsible for the PAM relaxation of UltraCjCas9. In CjCas9 WT, Thr913 forms a water-mediated hydrogen bond with the fourth PAM nucleotide and determines the specificity for V (A, C, or G) at this position due to the steric incompatibility of its side chain with the methyl group of thymine [41]. Mutation T913S likely avoids this steric clash thanks to the smaller side chain of serine compared to threonine. While structural considerations can explain the effect of this mutation on PAM preference, the structural basis for its effect on activity remains unclear. Interestingly, modifications introduced in UltraCjCas9 did not interfere with the nuclease specificity, as opposed to previous molecular engineering approaches [39], implying that enhanced nuclease activity does not necessarily increase tolerance to sgRNA-target DNA mismatches for Cas9 proteins, as also previously observed for CasΦ [46].

UltraCjCas9, obtained by directed evolution in EPICA, has attractive features for genome editing applications, including high editing efficiency (similar to SpCas9 at

specific genomic sites) and high specificity. Interestingly, while its editing activity is higher at most loci compared to CjCas9 WT, there are specific loci showing no increase in activity (e.g., *VEGFA2*, *HBG1*) or even a decrease (*VEGFA3*). This variability in editing activity is not uncommon for enhanced Cas nucleases [15, 47, 48] and could be ascribed to modified preferences for target sequence composition, which is known to affect Cas activity [49]. Finally, the reduced molecular size of UltraCjCas9 is compatible with the packaging capacity of AAV vectors, the most commonly used delivery method for genome editing [50, 51], consenting its delivery all-in-one with the sgRNA, as opposed to double AAVs for sgRNA and Cas9 for in vivo use [52–55].

Conclusions

Here, we report a fully eukaryotic platform for Cas enhancement. Our methodology exploits evolutionary steps in yeast followed by a selection of variants in mammalian cells. Enhanced nucleases such as UltraCjCas9, obtained through EPICA, will foster the expansion of the genome editing toolbox with the development of a larger variety of CRISPR-Cas systems. Since the platform allows the evolution of enzymes with nuclease activity, its use can be potentially extended beyond CRISPR-Cas systems, including the novel emerging collection of RNA-guided nucleases associated with prokaryotic and eukaryotic transposons [3–5, 56]. The modification of Cas9 proteins steered under the pressure imposed by the complexity of the nuclear environment may shed light on critical residues required for nuclease enhancement in eukaryotes, thus facilitating the translation of prokaryotic resources for genome editing applications.

Materials and methods

Plasmids

The constructs used for the generation of the yeast strain were obtained by cloning the AAVS1-TS34 and AAVS1-TS32 targets and PAM sequences as annealed oligonucleotides in pUC19 plasmid containing respectively ADE and TRP cassettes, after digestion with KpnI and BamHI. For expression in yeast, the sequence of CjCas9 was amplified from the pX404 (Addgene 68338) and cloned in place of SpCas9 in the LEU2 carrying plasmid p415-GalL-Cas9-CYC1t (Addgene 43804), by double digestion with SpeI/XhoI. pSNR52-BsmBI-Cj_sgRNA for sgRNA expression in yeast was generated by substituting the promoter in the pU6 Cj sgRNA plasmid (Addgene 89753) with the SNR promoter by double digestion with XhoI and BamHI. The AAVS1-TS34 and AAVS1-TS32 target sequences were cloned as annealed oligonucleotides in the pSNR52 Cj sgRNA digested with BsmBI. The SNR – AAVS1-TS34 and – TS32 Cj sgRNA fragments were first cloned in the URA3 carrying pRS316 plasmid by double digestion with XhoI/SacII adding the SUP4t yeast terminator; for double guide plasmid the SNR – AAVS1-TS32 Cj sgRNA fragment was then amplified and cloned in the generated pRS316-SNR52p-AAVS1-TS34sgRNA – SUP4t plasmid by digestion with SacI. The Sp sgRNA plasmids were obtained through PCR site-directed mutagenesis of p426-SNR52p-gRNA.CAN1.Y-SUP4t (Addgene 43803) to introduce the target sequences. For mammalian expression, the CjCas9 sgRNA scaffold was optimized through site-directed mutagenesis to interrupt the A/T stretches [33]; targets were cloned as annealed oligos in the pU6 opt CjCas9 sgRNA cut with BsmBI. The list of sgRNAs cloning primers is available in Additional

file 2: Table S4. For expression in mammalian cells, CjCas9 WT was cloned in a pX330 containing the chicken β -actin promoter by double digestion with AgeI and EcoRI/BbsI. To generate single mutants, CjCas9 wild-type sequence was cloned in a pEGFP-N1 with AgeI and MfeI and single substitutions introduced through PCR site-directed mutagenesis; the protein sequence was then cloned back in pX330. Multiple mutants were amplified from the two rounds enriched sorted cells and cloned in pX330 downstream of a SV40 NLS sequence previously inserted in the plasmid, as well as CjCas9 WT, digesting with KpnI and EcoRI/BbsI. enCjCas9 was generated as a double mutant introducing L58Y and D900K mutations in the CjCas9 WT sequence through site-directed mutagenesis and then cloned in the pX330 NLS plasmid. Primers used for plasmid cloning are reported in Additional file 2: Table S5.

Yeast culture

Yeast strains were grown in a YPDA-rich medium; for auxotrophic selection, yeasts were grown in a synthetic minimal medium (SD), omitting the single amino acids required for the experimental purpose. Yeasts stably expressing CjCas9 sgRNA were kept growing in SD without uracil before transformation. Cas9 expression was obtained using 20 g/L D-(+)-galactose and 10 g/L D-(+)-raffinose instead of dextrose.

Cell culture

HEK293 and HEK293T from the American Type Culture Collection (ATCC) were cultured in Dulbecco's modified Eagle's medium (DMEM; Life Technologies) with 10% FCS (Life Technologies) and antibiotics (Life Technologies). The EGFP T-REx 293 cell line used for the mammalian screening was obtained by modifying the Flp-InTM T-RExTM 293 Cell Line (Invitrogen) for the expression of inducible EGFP according to the manufacturer's protocol. The cell lines were verified for the absence of mycoplasma contamination (PlasmoTest, Invitrogen).

Yeast screening

The yeast strain used for the screening was generated from the γ LFM-ICORE strain by the Delitto Perfetto approach [57], using the constructs pUC19-Ade2-AAVS1TS34 and pUC19-Trp1-AAVS1TS32. The library of CjCas9 variants was produced by random mutagenesis of the wild-type sequence using error-prone PCR approach (GeneMorph II kit from Agilent) (for primers, see Additional file 2: Table S5). The manufacturer's conditions were followed to set a rate of 0–4.5 mutations/kb; yeasts were co-transformed with the PCR library and the previously digested (SpeI/XhoI) p415-GalL-CYC1t plasmid in a 3:1 ratio, to obtain direct assembly in vivo. Yeasts were transformed following the Gietz et al. published protocol [58]. After transformation, yeasts were grown for 5 h in an SD medium lacking uracil and leucine for recovery and recombination, then the expression of Cas9 was induced through overnight incubation in a galactose-containing medium. The next day, yeasts were plated in a medium lacking adenine and tryptophan, to select the colonies with enhanced variants. Colonies were collected, and the plasmids were extracted with Zymoprep Yeast Plasmid Miniprep II (Zymo Research). For the next rounds of mutagenesis, high-fidelity PCR was performed on the selected variants with PhusionTM High-Fidelity DNA Polymerase (Thermo Scientific) and used as a template for

the subsequent step of mutagenesis (described above). For each round of mutagenesis, transformation and collection of selected colonies were repeated. For the validation of the yeast strain, cells were transformed with the entire p415 plasmid expressing wild-type SpCas9 or CjCas9, using the same conditions of the screening. Yeast colonies were counted using the ImageJ software, manually setting the threshold for the maximum discrimination of the colonies. Each yeast experiment was performed by plating a defined amount of yeast suspension in $n = 3$ selective plates as technical replicates.

Deep sequencing

Variants of the fourth round of mutagenesis were PCR amplified with Phusion High-Fidelity DNA Polymerase (for primers, see Additional file 2: Table S5); the amplicon was used to generate a library with Nextera DNA Flex Library Preparation Kit (Illumina) and sequenced in a PE150 run using Illumina Miseq Reagent kit V2. For data analysis, reads were trimmed using TrimGalore [59] version 0.6.6 (with parameters `-q 30 -paired`) and aligned to the CjCas9 coding sequence using bowtie2 [60] version 2.3.4.3. Pileups were generated using PaCBAM [61] version 1.6.0 and a custom Python script was used to analyze mutation frequencies.

Indels analysis at human genomic loci

For the evaluation of genome editing activity, 10^5 HEK293 cells were seeded in 24-well plates and transfected the next day with 200 ng of sgRNA plasmid and 400 ng of Cas9 plasmid; TransIT-LT1 (Mirus Bio) was used as transfection reagent according to manufacturer's protocol. Cells were collected after 3 days and lysed with QuickExtract DNA Extraction Solution (Lucigen): endogenous loci were PCR amplified with HOT FIREPol MultiPlex Mix (Solis BioDyne), Sanger sequenced and analyzed with the TIDE tool [62]. The list of TIDE analysis primers is available in Additional file 2: Table S6.

Lentiviral vector library construction

The CjCas9 sgRNA targeting the TetR repressor was amplified from the pU6 plasmid and cloned in the LentiCRISPR V1 by double digestion with NdeI and EcoRI. LentiCRISPR V1 sgRNA TetR was then modified to have an intron upstream of the cloning site of Cas9, to avoid a leaky expression of the CjCas9 variants in bacteria and the potential loss of clones. An intron previously used for the same purpose [63] was cloned by insertional mutagenesis through PCR in the middle of the CjCas9 NLS sequence, between a splicing donor site and a splicing acceptor site: the construct Kozak-NLS-V5-intron was then amplified adding a KpnI site downstream and was cloned in LentiCRISPR using the restriction enzymes BamHI and NheI. To minimize the number of background colonies with undigested plasmid, the bacterial toxin *ccdB* was cloned with the enzymes KpnI and NheI using a dedicated commercial *ccdB*-resistant strain (Thermo Fisher); the generated plasmid LentiCRISPRV1-gRNA-TetR-intron-*ccdB* was used to obtain the KpnI/NheI digested backbone for lentiviral library. CjCas9 variants were amplified from the plasmid extract of the fourth round of the yeast screening, digested with KpnI and NheI and ligated at room temperature for 2 h with the digested backbone. The ligation was purified by NucleoSpin Gel and PCR Clean-up (Macherey–Nagel) and used for the electroporation of ElectroMAX

DH5 α -E Competent Cells (Invitrogen): bacterial colonies were collected until a coverage of at least 100 \times was reached. The number of different variants was estimated to be 1398, based on the total number of colonies collected from the fourth round of the yeast screening. Pooled library plasmids were purified using the NucleoBond PC 500 Maxi kit (Macherey–Nagel). The same protocol was applied to generate the library for the second round of selection, amplifying the variants from the genomic DNA of first-round sorted cells. Primers for the amplification were lengthened for each library to avoid the carry-over of unspecific products from the previous cloning. Primers used for lentiviral vector cloning are reported in Additional file 2: Table S5.

Cell transduction and sorting

For the lentiviral library production, 2×10^7 HEK293T cells (American Type Culture Collection) were transfected with 25 μ g of LentiCRISPR sgRNA TetR intron CjCas9 library, 16.2 μ g of pCMV-deltaR8.91 and 8.7 μ g of VSV-G using the polyethylenimine (PEI) method. Lentiviral particles were filtered through a 0.45- μ m PES filter and concentrated with ultracentrifugation. Vectors were resuspended in Opti-MEM (Gibco) and conserved at -80 °C. MOI was estimated by transducing a defined cell number with growing volumes of lentiviral particles and evaluating their viability in puromycin (1 μ g/m) after 2 days using an MTT assay. For the rounds of sorting, EGFP T-REx 293 cells were seeded and the next day were transduced with the vector library at MOI \sim 0.3, ensuring a $> 100 \times$ coverage of the library variants. Cells were then selected in puromycin (1 μ g/ml) 2 days after transduction and sorted for their EGFP fluorescence after 6 additional days using FACS ARIA III (BD Biosciences). Cells were grown for recovery and collected for genomic DNA extraction with a DNeasy Blood and Tissue kit (Qiagen).

In vitro PAM determination assay

The PAM identification assay was performed through in vitro cleavage of a randomized PAM library, adapting a previously established protocol (Altae-Tran et al. 2021). In brief, CjCas9 and UltraCjCas9 were in vitro transcribed and translated using PURExpress In vitro protein synthesis kit (NEB), along with their sgRNA targeting the PAM library spacer. Primers used are reported in Additional file 2: Table S5. In vitro cleavage was performed at the same time, adding 25 ng of randomized PAM library, previously created in our lab, in the mix and incubating the reaction for 4 h at 37°. The reaction was stopped by placing it on ice for 10 min and adding RNase A/T1 (Thermo Fisher) and Proteinase K (Thermo Fisher) to remove RNA and proteins. DNA was purified using 0.7 \times NGSclean beads (CleanNA): the ends were filled-in using T4 DNA polymerase (Thermo Fisher) and dA-tailed using DreamTaq DNA polymerase (Thermo Fisher). dsDNA oligo adapters were ligated to the fragments and a first PCR was performed as previously described [9]. The fragments of 200 bp were recovered from a size-selection purification with 0.5X NGSclean beads (CleanNA), purified with 1X NGSclean beads, and used as a template for a second round of PCR to attach Illumina indexes (NEBNext Multiplex Oligos for Illumina). Deep sequencing and data analysis were performed as previously described [9].

GUIDE-Seq

GUIDE-Seq was performed following a previously published protocol with minor modifications [25]. HEK293T cells (ATCC) were transfected with 250 ng of sgRNA plasmid, 500 ng of Cas9 plasmid, and 10 pmol of dsODNs, using Lipofectamine 3000 (Invitrogen). Cells were selected in puromycin (1 µg/ml) the next day, and after 4 days, genomic DNA was extracted using a DNeasy Blood and Tissue kit (Qiagen). Genomic DNA was sonicated to an average size of 500 bp using an S2 Focused-ultrasonicator (Covaris) and end-repaired with NEBNext Ultra End Repair/dA Tailing Module. Adaptors were ligated using the NEBNext Ultra Ligation Module as previously described [64]. After PCR amplifications, libraries were quantified by Qubit dsDNA High Sensitivity Assay kit (Invitrogen) and sequenced with the MiSeq sequencing system (Illumina) using Miseq Reagent kit V2-150PE.

CjCas9 structure

The crystal structure of CjCas9 WT in complex with its sgRNA and target DNA was retrieved from PDB (accession 5X2H) [41]. To investigate mutations in missing regions, the AlphaFold [65] predicted structure was retrieved from the AlphaFold Protein Structure Database (accession Q0P897) [44] and was aligned to the crystal structure using the Dali server [66]. Protein structures were visualized using ChimeraX [67].

Statistical analysis

Statistical significance tests were performed using GraphPad Prism (version 9.4.1). Data in Figs. 2b and 3d were analyzed with a one-way ANOVA followed by Tukey's multiple comparisons test. Data in Fig. 3e were analyzed using two-sided *t*-tests corrected for multiple comparisons using the Holm-Šidák method. For all analyses, adjusted *p*-values less than 0.05 were considered statistically significant.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13059-024-03215-9>.

Additional file 1: Fig. S1. Yeast cultures from the auxotrophic evolution platform. **Fig. S2.** TetR-EGFP reporter screening platform. **Fig. S3.** Editing activity of selected variants in endogenous loci of HEK293 cells. **Fig. S4.** 3D position of mutations in UltraCjCas9. **Fig. S5.** In vitro PAM characterization of CjCas9 WT and UltraCjCas9. **Fig. S6.** List of PAM sequences and editing at GUIDE-Seq ON-target sites. **Fig. S7.** Editing activity of CjCas9 WT and UltraCjCas9 in the *PDCD1* locus using mismatched sgRNAs.

Additional file 2: Table S1. Indel frequencies of selected variants. **Table S2.** Sanger sequences of selected variants. **Table S3.** Indel frequencies on genomic loci. **Table S4.** sgRNA cloning primers. **Table S5.** Cloning primers. **Table S6.** TIDE analysis primers. **Table S7.** Identified off-targets of CjCas9 WT targeting AAVS1TS1. **Table S8.** Identified off-targets of UltraCjCas9 targeting AAVS1TS1. **Table S9.** Identified off-targets of CjCas9 WT targeting AAVS1TS8. **Table S10.** Identified off-targets of UltraCjCas9 targeting AAVS1TS8. **Table S11.** Indel frequencies with mismatched sgRNAs.

Additional file 3. Review history.

Acknowledgements

We are grateful to Cereseto's lab for the helpful discussion throughout the project. We thank the next-generation sequencing facility at the University of Trento for the technical support.

Peer review information

Kevin Pang was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Review history

The review history is available as Additional file 3.

Authors' contributions

G.V.R., E.K., M.D.G., and S.A. designed and performed the experiments. M.C. performed the computational analyses and analyzed the data. G.V.R., M.C., A.Ca., and A.Ce. wrote and edited the paper. A.Ce. and A.Ca. conceived and designed the study. A.Ce. was responsible for the coordination of the study. All authors read, corrected, and approved the final manuscript.

Funding

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) and Horizon Europe EIC Pathfinder Program AAVolution (grant agreement 01071041).

Availability of data and materials

Deep sequencing data of the CJCas9 library derived from the 4th round of mutagenesis and GUIDE-Seq data for CJCast WT and UltraCjCas9 are available on NCBI (BioProject PRJNA1020659) [68]. The custom code used in this study is available on Zenodo and GitHub: <https://github.com/Matteo-Ciciani/EPICA> [69, 70].

Declarations**Ethics approval and consent to participate.**

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare competing financial interests: A.Ce. is a co-founder and holds shares of Alia Therapeutics, a genome editing company. A.Ca. is a co-founder, holds shares, and is currently an employee of Alia Therapeutics. M.C. is a consultant of Alia Therapeutics.

Received: 26 September 2023 Accepted: 13 March 2024

Published online: 25 March 2024

References

- Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, Bryson DI, et al. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature*. 2017;551:464–71.
- Anzalone AV, Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*. 2019;576:149–57.
- Altae-Tran H, Kannan S, Demircioglu FE, Oshiro R, Nety SP, McKay LJ, et al. The widespread IS200/IS605 transposon family encodes diverse programmable RNA-guided endonucleases. *Science*. 2021;374:57–65.
- Karvelis T, Druteika G, Bigelyte G, Budre K, Zedaveinyte R, Silanskas A, et al. Transposon-associated TnpB is a programmable RNA-guided DNA endonuclease. *Nature*. 2021;599:692–6.
- Saito M, Xu P, Faure G, Maguire S, Kannan S, Altae-Tran H, et al. Fanzor is a eukaryotic programmable RNA-guided endonuclease. *Nature*. 2023; Available from: <https://doi.org/10.1038/s41586-023-06356-2>
- Doudna JA. The promise and challenge of therapeutic genome editing. *Nature*. 2020;578:229–36.
- Pickar-Oliver A, Gersbach CA. The next generation of CRISPR-Cas technologies and applications. *Nat Rev Mol Cell Biol*. 2019;20:490–507.
- Mojica FJM, Díez-Villaseñor C, García-Martínez J, Almendros C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*. 2009;155:733–40.
- Ciciani M, Demozzi M, Pedrazzoli E, Visentin E, Pezzè L, Signorini LF, et al. Automated identification of sequence-tailored Cas9 proteins using massive metagenomic data. *Nat Commun*. 2022;13:6474.
- Davis JR, Wang X, Witte IP, Huang TP, Levy JM, Raguram A, et al. Efficient in vivo base editing via single adeno-associated viruses with size-optimized genomes encoding compact adenine base editors. *Nat Biomed Eng*. 2022;6:1272–83.
- Wang D, Zhang F, Gao G. CRISPR-based therapeutic genome editing: strategies and in vivo delivery by AAV vectors. *Cell*. 2020;181:136–50.
- Gasiunas G, Young JK, Karvelis T, Kazlauskas D, Urbaitis T, Jasnauskaitė M, et al. A catalogue of biochemically diverse CRISPR-Cas9 orthologs. *Nat Commun*. 2020;11:5512.
- Makarova KS, Wolf YI, Iranzo J, Shmakov SA, Alkhnbashi OS, Brouns SJJ, et al. Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nat Rev Microbiol*. 2020;18:67–83.
- Al-Shayeb B, Skopintsev P, Soczek KM, Stahl EC, Li Z, Groover E, et al. Diverse virus-encoded CRISPR-Cas systems include streamlined genome editors. *Cell*. 2022;185:4574–86.e16.
- Nakagawa R, Ishiguro S, Okazaki S, Mori H, Tanaka M, Aburatani H, et al. Engineered campylobacter jejuni Cas9 variant with enhanced activity and broader targeting range. *Commun Biol*. 2022;5:211.
- Zhang L, Zuris JA, Viswanathan R, Edelstein JN, Turk R, Thommandru B, et al. Author correction: AsCas12a ultra nucle-ase facilitates the rapid generation of therapeutic cell medicines. *Nat Commun*. 2021;12:4500.

17. Kim DY, Lee JM, Moon SB, Chin HJ, Park S, Lim Y, et al. Efficient CRISPR editing with a hypercompact Cas12f1 and engineered guide RNAs delivered by adeno-associated virus. *Nat Biotechnol.* 2022;40:94–102.
18. Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, Kriz AJ, et al. In vivo genome editing using staphylococcus aureus Cas9. *Nature.* 2015;520:186–91.
19. Casini A, Olivieri M, Petris G, Montagna C, Reginato G, Maule G, et al. A highly specific SpCas9 variant is identified by in vivo screening in yeast. *Nat Biotechnol.* 2018;36:265–71.
20. Kim E, Koo T, Park SW, Kim D, Kim K, Cho H-Y, et al. In vivo genome editing with a small Cas9 orthologue derived from *Campylobacter jejuni*. *Nat Commun.* 2017;8:14500.
21. Li F, Wing K, Wang J-H, Luu CD, Bender JA, Chen J, et al. Comparison of CRISPR/Cas endonucleases for in vivo retinal gene editing. *Front Cell Neurosci.* 2020;14:570917.
22. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. *Science.* 2015;347:1260419.
23. Schmid-Burgk JL, Gao L, Li D, Gardner Z, Strecker J, Lash B, et al. Highly parallel profiling of Cas9 variant specificity. *Mol Cell.* 2020;78:794–800.e8.
24. Walton RT, Christie KA, Whittaker MN, Kleinstiver BP. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science.* 2020;368:290–6.
25. Tsai SQ, Zheng Z, Nguyen NT, Liebers M, Topkar VV, Thapar V, et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol.* 2015;33:187–97.
26. Daer RM, Cutts JP, Brafman DA, Haynes KA. The impact of chromatin dynamics on Cas9-mediated genome editing in human cells. *ACS Synth Biol.* 2017;6:428–38.
27. Wu X, Scott DA, Kriz AJ, Chiu AC, Hsu PD, Dadon DB, et al. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat Biotechnol.* 2014;32:670–6.
28. Chen X, Rinsma M, Janssen JM, Liu J, Maggio I, Gonçalves MAFV. Probing the impact of chromatin conformation on genome editing tools. *Nucleic Acids Res.* 2016;44:6482–92.
29. Yarrington RM, Verma S, Schwartz S, Trautman JK, Carroll D. Nucleosomes inhibit target cleavage by CRISPR-Cas9 in vivo. *Proc Natl Acad Sci U S A.* 2018;115:9351–8.
30. Isaac RS, Jiang F, Doudna JA, Lim WA, Narlikar GJ, Almeida R. Nucleosome breathing and remodeling constrain CRISPR-Cas9 function. *Elife.* 2016;5:e13450. <https://doi.org/10.7554/eLife.13450>.
31. Horlbeck MA, Witkowsky LB, Guglielmi B, Replogle JM, Gilbert LA, Villalta JE, et al. Nucleosomes impede Cas9 access to DNA in vivo and in vitro. *Elife.* 2016;5:e12677. <https://doi.org/10.7554/eLife.12677>.
32. Strohkendl I, Saifuddin FA, Gibson BA, Rosen MK, Russell R, Finkelstein JI. Inhibition of CRISPR-Cas12a DNA targeting by nucleosomes and chromatin. *Sci Adv.* 2021;7:eabd6030. <https://doi.org/10.1126/sciadv.abd6030>.
33. Ding X, Seebeck T, Feng Y, Jiang Y, Davis GD, Chen F. Improving CRISPR-Cas9 genome editing efficiency by fusion with chromatin-modulating peptides. *CRISPR J.* 2019;2:51–63.
34. Dong L, Guan X, Li N, Zhang F, Zhu Y, Ren K, et al. An anti-CRISPR protein disables type V Cas12a by acetylation. *Nat Struct Mol Biol.* 2019;26:308–14.
35. Niu Y, Yang L, Gao T, Dong C, Zhang B, Yin P, et al. A Type I-F Anti-CRISPR protein inhibits the CRISPR-Cas surveillance complex by ADP-ribosylation. *Mol Cell.* 2020;80:512–24.e5.
36. Kang X, Yin L, Zhuang S, Hu T, Wu Z, Zhao G, et al. Reversible regulation of Cas12a activities by AcrVA5-mediated acetylation and CobB-mediated deacetylation. *Cell Discov.* 2022;8:45.
37. Ergünay T, Ayhan Ö, Celen AB, Georgiadou P, Pekbilir E, Abaci YT, et al. Sumoylation of Cas9 at lysine 848 regulates protein stability and DNA binding. *Life Sci Alliance.* 2022;5. <https://doi.org/10.26508/lsa.202101078>
38. Kleinstiver BP, Prew MS, Tsai SQ, Topkar VV, Nguyen NT, Zheng Z, et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature.* 2015;523:481–5.
39. Kleinstiver BP, Sousa AA, Walton RT, Tak YE, Hsu JY, Clement K, et al. Engineered CRISPR-Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. *Nat Biotechnol.* 2019;37:276–82.
40. Slaymaker IM, Gao L, Zetsche B, Scott DA, Yan WX, Zhang F. Rationally engineered Cas9 nucleases with improved specificity. *Science.* 2016;351:84–8.
41. Yamada M, Watanabe Y, Gootenberg JS, Hirano H, Ran FA, Nakane T, et al. Crystal structure of the minimal Cas9 from *Campylobacter jejuni* reveals the molecular diversity in the CRISPR-Cas9 systems. *Mol Cell.* 2017;65:1109–21.e3.
42. Hirano S, Abudayyeh OO, Gootenberg JS, Horii T, Ishitani R, Hatada I, et al. Structural basis for the promiscuous PAM recognition by *Corynebacterium diphtheriae* Cas9. *Nat Commun.* 2019;10:1968.
43. Das A, Hand TH, Smith CL, Wickline E, Zawrotny M, Li H. The molecular basis for recognition of 5'-NNNCC-3' PAM and its methylation state by *Acidothermus cellulolyticus* Cas9. *Nat Commun.* 2020;11:6346.
44. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* 2022;50:D439–44.
45. Pacesa M, Loeffl L, Querques I, Muckenfuss LM, Sawicka M, Jinek M. R-loop formation and conformational activation mechanisms of Cas9. *Nature.* 2022;609:191–6.
46. Pausch P, Soczek KM, Herbst DA, Tsuchida CA, Al-Shayeb B, Banfield JF, et al. DNA interference states of the hypercompact CRISPR-CasΦ effector. *Nat Struct Mol Biol.* 2021;28:652–61.
47. Xu X, Chemparathy A, Zeng L, Kempton HR, Shang S, Nakamura M, et al. Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing. *Mol Cell.* 2021;81:4333–45.e4.
48. Hino T, Omura SN, Nakagawa R, Togashi T, Takeda SN, Hiramoto T, et al. An AsCas12f-based compact genome-editing tool derived by deep mutational scanning and structural analysis. *Cell.* 2023. <https://doi.org/10.1016/j.cell.2023.08.031>
49. Kim HK, Song M, Lee J, Menon AV, Jung S, Kang Y-M, et al. In vivo high-throughput profiling of CRISPR-Cpf1 activity. *Nat Methods.* 2017;14:153–9.
50. van Haasteren J, Li J, Scheideler OJ, Murthy N, Schaffer DV. The delivery challenge: fulfilling the promise of therapeutic genome editing. *Nat Biotechnol.* 2020;38:845–55.

51. Pupo A, Fernández A, Low SH, François A, Suárez-Amarán L, Samulski RJ. AAV vectors: the Rubik's cube of human gene therapy. *Mol Ther.* 2022;30:3515–41.
52. Levy JM, Yeh W-H, Pendse N, Davis JR, Hennessey E, Butcher R, et al. Cytosine and adenine base editing of the brain, liver, retina, heart and skeletal muscle of mice via adeno-associated viruses. *Nat Biomed Eng.* 2020;4:97–110.
53. Villiger L, Grisch-Chan HM, Lindsay H, Ringnalda F, Pogliano CB, Allegri G, et al. Treatment of a metabolic liver disease by in vivo genome base editing in adult mice. *Nat Med.* 2018;24:1519–25.
54. Davis JR, Banskota S, Levy JM, Newby GA, Wang X, Anzalone AV, et al. Efficient prime editing in mouse brain, liver and heart with dual AAVs. *Nat Biotechnol.* 2023. <https://doi.org/10.1038/s41587-023-01758-z>
55. Qin H, Zhang W, Zhang S, Feng Y, Xu W, Qi J, et al. Vision rescue via unconstrained in vivo prime editing in degenerating neural retinas. *J Exp Med.* 2023;220:e20220776. <https://doi.org/10.1084/jem.20220776>.
56. Xiang G, Li Y, Sun J, Huo Y, Cao S, Cao Y, et al. Evolutionary mining and functional characterization of TnpB nucleases identify efficient miniature genome editors. *Nat Biotechnol.* 2023. <https://doi.org/10.1038/s41587-023-01857-x>
57. Storici F, Resnick MA. The delitto perfetto approach to in vivo site-directed mutagenesis and chromosome rearrangements with synthetic oligonucleotides in yeast. *Methods Enzymol.* 2006;409:329–45.
58. Gietz RD, Schiestl RH. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc.* 2007;2:31–4.
59. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 2011;17:10.
60. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9:357–9.
61. Valentini S, Fedrizzi T, Demichelis F, Romanel A. PaCBAM: fast and scalable processing of whole exome and targeted sequencing data. *BMC Genomics.* 2019;20:1018.
62. Brinkman EK, Chen T, Amendola M, van Steensel B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* 2014;42: e168.
63. Petris G, Casini A, Montagna C, Lorenzin F, Prandi D, Romanel A, et al. Hit and go CAS9 delivered through a lentiviral based self-limiting circuit. *Nat Commun.* 2017;8:15334.
64. Nobles CL, Reddy S, Salas-McKee J, Liu X, June CH, Melenhorst JJ, et al. iGUIDE: an improved pipeline for analyzing CRISPR cleavage specificity. *Genome Biol.* 2019;20:14.
65. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature.* 2021;596:583–9.
66. Holm L. Dali server: structural unification of protein families. *Nucleic Acids Res.* 2022;50:W210–5.
67. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, et al. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* 2021;30:70–82.
68. Ruta GV, Ciciani M, KheirE, Gentile MD, Amistadi S, Casini A, et al. Eukaryotic-driven directed evolution of Cas9 nucleases. *NCBI SRA.* <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA1020659> (2023).
69. Ciciani M. EPICA Code 1.0.0. Zenodo. <https://doi.org/10.5281/zenodo.10804018>.
70. Ciciani M. EPICA Code 1.0.0. Github. <https://github.com/Matteo-Ciciani/EPICA>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.