Doctoral School of Multidisciplinary Medical Science
Faculty of Medicine
University of Szeged

Development of new increased fidelity SpCas9 variants

Péter István Kulcsár

PhD Thesis

Supervisor:
Ervin Welker PhD, DSc

Institute of Biochemistry
Biological Research Centre
Hungarian Academy of Sciences

Institute of Enzymology
Research Centre for Natural Sciences
Hungarian Academy of Sciences

Szeged
2019

List of Publications

Publications related to the thesis:

I. **Kulcsár P. I.**, Tálas A., Huszár K., Ligeti Z., Tóth E., Weinhardt N., Fodor E., Welker E., Crossing enhanced and high fidelity SpCas9 nucleases to optimize specificity and cleavage, Genome Biology (2017). **IF: 13.214**

II. Tálas A., **Kulcsár P. I.**, Weinhardt N., Borsy A., Tóth E., Szebényi K., Krausz S. L., Huszár K., Vida I., Sturm Á., Gordos B., Hoffmann O. I., Bencsura P., Nyeste A., Ligeti Z., Fodor E., Welker E., A convenient method to pre-screen candidate guide RNAs for CRISPR/Cas9 gene editing by NHEJ-mediated integration of a 'self-cleaving' GFP-expression plasmid, DNA Research (2017). **IF: 5.415**

III. **Kulcsár P. I.**, Tálas A., Tóth E., Ligeti Z., Welker Zs., Welker E., Blackjack mutations improve the on-target activities of all increased fidelity variants of SpCas9 without compromising their fidelities (under review).

Cumulative impact factor of papers directly related to the thesis: 18.629

Publications indirectly related to the subject of the thesis:

IV. Tóth E., Czene B. C., **Kulcsár P. I.**, Krausz S. L., Tálas A., Nyeste A., Varga É., Huszár K., Weinhardt N., Ligeti Z., Borsy A. É., Fodor E., Welker E., Mb-and FnCpf1 nucleases are active in mammalian cells: activities and PAM preferences of four wild-type Cpf1 nucleases and of their altered PAM specificity variants, Nucleic acids research (2018). **IF: 11.147**

V. Tóth E., Weinhardt N., Bencsura P., Huszár K., **Kulcsár P. I.**, Tálas A., Fodor E., Welker E., Cpf1 nucleases demonstrate robust activity to induce DNA modification by exploiting homology directed repair pathways in mammalian cells, Biology Direct (2016). **IF: 2.856**

Publications not related to the thesis:

VI. Billes V., Kovács T., Manzéger A., Lőrincz P., Szincsák S., Rezgős Á., **Kulcsár P. I.**, Korcsmáros T., Lukácsovich T., Hoffmann Gy., Erdélyi M., Mihály J.,

Takács-Vellai K., Sass M., Vellai T., Developmentally regulated autophagy is required for eye formation in *Drosophila*, Autophagy (2018). **IF: 11.059**

VII. Tóth E., Huszár K., Bencsura P., **Kulcsár P. I.**, Vodicska B., Nyeste A., Welker Zs., Tóth Sz., Welker E., Restriction Enzyme Body Doubles and PCR Cloning: On the General Use of Type IIS Restriction Enzymes for Cloning, Plos One (2014). **IF: 3.234**

VIII. Kulcsár G., Gaál D., **Kulcsár P.I.**, Schulcz Á., Czömpöly T. A mixture of amino acids and other small molecules present in the serum suppresses the growth of murine and human tumors in vivo., Int J Cancer, (2013). **IF: 5.007**

IX. Tóth E., **Kulcsár P. I.**, Fodor E., Ayaydin F., Kalmár L., Borsy A. E., László L., Welker E., The highly conserved, N-terminal (RXXX)(8) motif of mouse Shadoo mediates nuclear accumulation., Biochim Biophys Acta (2013). **IF: 5.297**

Cumulative impact factor of all papers: 57.229, h-index: 5, citations: 102

Author's contribution

My PhD thesis is primarily based on three publications (one of them is under review) listed on the earlier pages. All experiments shown here were designed and performed by me unless stated otherwise. Here – and also in every relevant figure legend and method section - I list the contributions of the co-authors.

- András Tálas performed the RNP experiment shown in Figure 14 and cloned 20 out of 430 plasmid constructs.

- Zoltán Ligeti did most of the *in vitro* transcriptions, the cloning of 7 out of 430 plasmid constructs, the protein expressions and the EMSA experiments shown in Figure 8.

- Krisztina Huszár performed part of the experiment shown in Figure 5 and cloned 60 out of 430 plasmid constructs.

- Elfrieda Fodor calculated the Pearson correlations shown in Figure 6.

Table of Contents

Abbreviations

B- prefix	Blackjack mutations (aa. 1005-1013 replaced in SpCas9 with two glycine)
Cas9	CRISPR associated protein 9
CMV	Human Cytomegalovirus
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
crRNA	CRISPR RNA
DSB	Double Strand Break
dSpCas9	dead/inactive SpCas9
EGFP	Enhanced Green Fluorescent Protein
EMSA	Electrophoretic Mobility Shift Assay
GUIDE-seq	Genome-wide, Unbiased Identification of DSBs Enabled by sequencing
HDR	Homology-Directed Repair
MN	Meganuclease
NGS	Next-Generation Sequencing
NHEJ	Non-Homologous End Joining
NLS	Nuclear Localization Signal
PAM	Protospacer-Adjacent Motif
RGN	RNA-Guided Endonuclease
RNP	Ribonucleoprotein
sgRNA	single guide RNA
SpCas9	*Streptococcus pyogenes* CRISPR associated protein 9
Sprn	Shadow of Prion protein
TALEN	Transcription Activator-Like Effector Nuclease
tracrRNA	trans-activating crRNA
WGS	Whole Genome Sequencing
WT	wild type
ZFN	Zinc Finger Nuclease

1. Introduction

The development of new genome modification techniques has greatly increased the feasibility of introducing targeted changes into the genome of any target organism. These techniques have great potential in medical treatments[1-6] , in agriculture[7-9] and in basic research applications as well[5,10-14]. To get genome engineering up to the current level, one of the biggest challenges needed to be addressed was to develop reliable methods for targeted DNA modification. Historically, the first genetic modifications relied on spontaneous mutations. Later, the mutagenesis rate was increased using chemical or radioactive agents, or by using transposons which worked with both low efficiency and low precision[15]. It has also been shown that cells can stably incorporate an exogenous DNA sequence at random positions into their genome, due to randomly occurring double strand break (DSB) events[16]. In some cell types, targeted integration was performed using homologous sequences, but with an extremely low rate[17]. All these genome modification approaches hijack the DNA repair pathways of the host cell.

The mechanism of DNA repair in eukaryotic cells after a DSB event can be classified into two major groups: non-homologous end joining (NHEJ) and homology-directed repair (HDR). The choice between these pathways is regulated at the initial step of the repair. NHEJ mediates the repair of two broken DNA ends by directly ligating them, therefore it does not need any template for the repair. NHEJ is the predominant repair pathway, although in some cases it might lead to error-prone repair, i.e. introducing insertions or deletions (indels) into the sequence. These indels could cause frameshift and/or early stops when occurring in the coding region of a protein which could lead to diminished protein expression[18,19]. By contrast, in case of HDR a homologous DNA sequence is needed. This pathway leads to more precise, error free repair, although it happens in the lower fraction of the events[19,20]. It has been shown that the introduction of targeted DSBs into the genome to initiate the repair pathways resulted in an increase of several orders of magnitude in the frequency of the desired mutation at the target site, therefore the search for nucleases capable of targeted DSB cleavage got into the focus[21,22].

The discovery of the restriction endonucleases was the first big step in the field of DNA modifications[23-25]. Although they are routinely used in molecular cloning techniques and *in vitro* applications even nowadays, they do not meet expectations in

eukaryotic cell modifications. The two major limitations of the restriction endonucleases are that their recognition sequences (i) cannot be changed (i.e. they are not modular), therefore the number of available target sites are limited and (ii) they are short, typically only 4-8 bp long. For eukaryotic cells, enzymes that recognize longer sequences - which are able to make unique targeted DSBs even in the context of a complex genome - are desired[26].

Meganucleases (MNs), or homing endonucleases, were the first in the line found by researchers with longer, 14-40 bp recognition sites. Their major limitation was also the low number of available target sequences, due to the lack of modularity[15,22,27]. The discovery of zinc finger nucleases (ZFNs) was a major breakthrough in the field. It is easier to redesign ZFN proteins to target a new genomic site compared to MNs due to their modular design, although it requires extensive protein engineering work. Zinc finger endonucleases are generated by combining the DNA cleavage domain of the *Fok*I restriction endonuclease and a series of small zinc-ion regulated protein domains, each bind to specific 3 bp-long DNA triplets. *Fok*I nuclease cleaves the DNA only in dimer form, therefore two separate ZFN have to be designed to target two genomic sites (on separate strands) next to each other which allows *Fok*I to homodimerize and introduce DSB at the desired site[27,28].

The discovery of transcription activator-like effector (TALE) proteins took the field one step further. TALEs recognize single bases instead of 3 bp motifs which makes it even easier to design new modular nucleases by fusing a series of DNA binding domains of TALE proteins and the DNA cleavage domain of *Fok*I. TALE nucleases (TALENs) are considered to have higher on-target activity, higher specificity, are less genotoxic and constructing them is easier than ZFNs[29]. All of these nucleases (MN, ZFN and TALEN) achieve the specific DNA binding by distinct protein-DNA interactions. Although they can bind the DNA and introduce the DSB with high specificity, they have some serious limitations such as the need for complex molecular cloning methods and extensive protein engineering, involving a lot of hands-on time required to clone the desired variants for each different chosen on-target site (Table 1)[15,27,30]. These limitations prevented ZFNs and TALENs from becoming broadly used tools by the scientific community, although there are several treatments under development that exploit ZFN or

TALEN[3,31]. ZFN has been tested in clinical trial for the treatment of HIV[32] and two children with advanced leukemia have been successfully treated with TALEN[33,34].

The introduction of the CRISPR (clustered regularly interspaced short palindromic repeats)/Cas9 (CRISPR associated protein 9) system as a genome engineering tool clearly revolutionized the field of biotechnology. CRISPR/Cas proteins are RNA-guided endonucleases (RGNs) that can be directed to cleave a chosen DNA[10,35,36] or RNA[37] sequence and they provide an adaptive immunity against mobile viruses, plasmids and transposons in archaea and bacteria[38,39]. The CRISPR/Cas systems can be classified into two major classes according to our current knowledge. Class 1 encompasses the type I, type III and type IV groups which have multiple subunit effector complexes. Class 2 contains much simpler systems with single multifunctional and multidomain protein effector modules. Class 2 consist of type II (including the Cas9 proteins), type V and type VI groups. Nucleases of type II systems contain cas proteins with similar domain architecture, including a RuvC-like and a HNH nuclease domains each cleaving one DNA strand. Type V (such as Cas12a proteins [former name: Cpf1]) systems contain effectors with only one active RuvC-like nuclease domain for cleaving both DNA strands, while in case of type VI subtypes two HEPN RNase domains are presented[40-42].

	ZFNs	**TALENs**	**CRISPR/Cas**
Type of target recognition	protein - DNA	protein - DNA	RNA - DNA
Key components	Zinc fingers fused to *Fok*I endonuclease	TALE domains fused to *Fok*I endonuclease	Cas protein, cr- and tracrRNA/sgRNA
Target site length (bp)	18-36	30-40	20-24
Retargeting/construction	Difficult to design High cost	Moderate design and cost Modular	Simple Easy to generate Low cost
Advantages	Easy in vivo delivery	High specificity	Good on-target efficiency (depends on the target) Easy multiplexing
Limitations	Possible cytotoxic off-target effect Not all DNA triplets have appropriate ZFs	Difficult in vivo delivery	PAM requirement Higher off-target effect (depends on the target) Moderate in vivo delivery

Table 1. Comparison of different engineered nucleases used for targeted gene editing. The table shows the advantages and limitations of zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and the CRISPR (clustered regularly interspaced short palindromic repeats)/Cas9 (CRISPR associated protein 9) system[15,26,27].

Since the first published results, many different CRISPR nucleases such as *Streptococcus pyogenes* Cas9 (SpCas9)[10,35,36,43,44], Streptococcus thermophilus Cas9[45], *Staphylococcus aureus* Cas9[46], *Neisseria meningitidis* Cas9[47]; *Acidaminococcus sp.* (strain BV3L6) Cas12a, *Lachnospiraceae bacterium MA2020* Cas12a, *Francisella novicida U112* Cas12a, *Moraxella bovoculi 237* Cas12a[48-51]; CasX and CasY[52,53] have proved their values for genome engineering applications, but among them presently SpCas9 is the most commonly used genome engineering tool[12,13,54-62].

The ribonucleoprotein (RNP) complex of the Cas9 nucleases involve the Cas9 protein itself and two Cas9-associated RNAs [*CRISPR* RNA (crRNA) and the *trans-activating* crRNA (tracrRNA)] possessing sequences complementary to each other. Complementarity between the targeted DNA site and the spacer sequence of the crRNA and the presence of a short protospacer-adjacent motif (PAM) at the 3'-end of the target site are also required to the binding and cleavage to occur[36,43,44,63-68]. The length of the required spacer sequence and the PAM motif varies depending on the species the Cas9 originates from. In case of SpCas9 a 20-nucleotide long spacer sequence and an NGG PAM motif downstream of the target sequence on the non-targeted DNA strand are needed[69] (Fig. 1A). It has been shown that SpCas9 nucleases can be guided to the desired target site by a fused cr- and tracrRNA, named *single guide* RNA (sgRNA, sometimes referred as gRNA)[43] (Fig. 1B). CRISPR/Cas nucleases show not only a comparable, or indeed higher level of on-target activity as ZFNs and TALENs, but it is much cheaper and simpler to redirect them to a new genomic target site, specifically, only the spacer sequence of the cr- or sgRNA has to be changed[15] (Table 1).

Although the use of SpCas9 has a lot of advantages compared to the earlier mentioned systems, some limitations still remain. The two major ones that have to be overcome are related to its varying efficacy (i.e. level of percentage modified at the desired on-target site) and specificity (i.e. on-target activity vs. off-target cleavage). There are further challenges such as delivery, pharmacokinetics, safety and immunogenicity which have to be resolved for effective clinical applications. SpCas9 nuclease shows high level of on-target cleavage in case of most of the target sites, but also a considerable level of off-target activity; i.e. the nuclease also cleaves targets that show limited, imperfect complementarities with the associated sgRNA[70-77] (Fig. 1C). The off-target sequences are difficult to predict *in silico* and have been shown to contain up to 5-6 base

mismatches[78,79], a property that may interfere with many research applications as well as using this nuclease for therapeutic purposes.

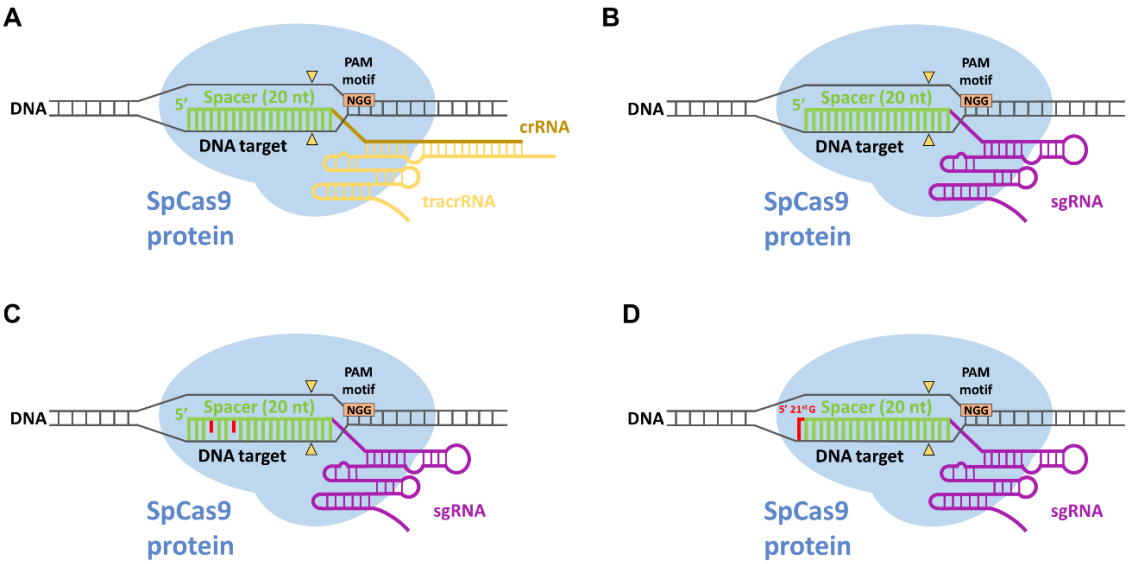


Figure 1. Schematics of SpCas9 nuclease complex with bound DNA target. A-D, SpCas9 nuclease (blue) targeted to genomic DNA (black) charged with (**A**) crRNA (brown) and tracrRNA (yellow), or (**B-D**) with sgRNAs (scaffold in purple): (**A-B**) perfectly matching 20G-sgRNA, (**C**) partially mismatching sgRNA or (**D**) 5' G extended 21G-sgRNA. The NGG PAM motif (orange) has to be directly downstream at the 3'-end of the target DNA sequence. SpCas9 mediates a DSB ~3 bp upstream of the PAM motif (yellow triangles). Spacers are schematically depicted as combs: green color teeth indicate matching-, while red color teeth indicate the presence of a mismatching nucleotide within the spacer.

There are different approaches developed for the identification of the truly cleaved off-target sites by SpCas9. One can PCR amplify the *in silico* predicted target sites and either perform a T7E1 or Surveyor assay[80] or sequence them. The sequencing could be done either by traditional Sanger sequencing (indel percentages from chromatograms can be calculated with algorithms such as TIDE[81]), or by next-generation sequencing (NGS). These methods are time and cost effective, but their main disadvantage is that not all off-target sites can be predicted, and not all predicted off-target sites are truly cleaved. It also has been proposed that whole-genome sequencing (WGS) could be an option for unbiased detection of real cleavage events. Although WGS is applicable for single-cell clone validation, it is not applicable in case of cell populations[82,83]. Since the spread of the

CIRSPR/Cas systems different unbiased methods have been developed for *in vitro* (Digenome-seq[84], SITE-seq[85], CIRCLE-seq[86]) and *in vivo* (IDLV[77,87], HTGTS[88], BLISS[89], GUIDEseq[75], DISCOVER-Seq[90]) identification of the off-target sites. Although there are currently numerous ways available to identify off-target sites cleaved by SpCas9, in some cell lines or *in vivo* it could be still a challenge to detect them. In addition, it is questionable whether the detection limit of these methods is sufficiently low enough to identify rare off-target events that still should be avoided for sensitive clinical applications. Experiments using these methods confirmed that wild type (WT) SpCas9 has significant off-target effect on many target sites, therefore developing CRISPR systems which help to overcome this limitation has paramount importance.

One of the approaches to facilitate the selection of appropriate targets (that could be cleaved efficiently without any off-targets) is the development of engineered SpCas9 mutants which increase the available target space by relaxing or changing the PAM preferences. An expanded target space hypothetically allows us to choose on-target sites with lower number of potential off-target sites at the desired genomic region. In case of VQR SpCas9 variants the protein has been engineered to recognize NGA, NGAG and NGCG PAM motifs[91]. SpCas9-NG[92] and xSpCas9[93] have been designed that way to recognize a broader range of PAMs, namely only an NG restriction has to be fulfilled instead of an NGG motif. These variants are greatly broadening the available targets space, although, on the other side they need to deal with a higher number of possible off-target sites due to their shorter PAM motifs.

Much effort has been devoted to circumventing the off-target effect of the SpCas9 nuclease with different approaches. There are two main directions for developing methods to attain higher specificity SpCas9 cleavage. (i) In case of the first direction one tries to limit the actions of SpCas9 in the cells such a way that it has still enough activity for on-target cleavage but no, or lower, activity for off-target cleavage. This is generally attempted by restricting the time or the extent of the exposure (e.g. applying inducible Cas9 or *in vitro* expressed ribonucleoprotein complex)[57,70,94-96] or by limiting the activity of the protein through the use of sgRNAs containing truncated[75,78,97] or extended spacer sequences[74,98]. (ii) The other direction is to increase the recognition sequence by using paired SpCas9 nickases[11,74,99] or using pairs of catalytically inactive SpCas9 fused to a non-specific *Fok*I nuclease domain[100-102].

The most promising approaches to decrease its off-target activity, however, are the generation of increased fidelity mutant variants. The first two published increased fidelity variants were eSpCas9[103] and SpCas9-HF1[104]. They were made by rational design: eSpCas9 was developed with the concept to decrease the affinity of the protein for the non-target DNA strand, thus increasing the strand's propensity for reinvading the RNA-DNA hybrid helix, and therefore decreasing the stability of mismatch-containing helices[103]. By contrast, mutations contained in the SpCas9-HF1 weaken the interactions of the protein with the target DNA strand aimed at decreasing the energetics of the SpCas9–sgRNA complex so that it retains a robust on-target activity but has a diminished ability to cleave mismatched off-target sites[104]. Both mutants exhibited considerably reduced off-target effects even when assessed by unbiased whole-genome off-target analysis (BLESS[105] and GUIDE-seq[75], respectively), although some off-targets, mainly only with single-base mismatches, were still found[103,104]. However, a subset of the targets, referred to as atypical in Kleinstiver et al., with repetitive or homopolymeric sequences were still cleaved with considerably high off-target effects[104]. Two additional increased fidelity variants were also published later: HypaSpCas9 developed by rational design as well[106] and evoSpCas9 by exploiting a selection scheme[107].

It has been shown that increased fidelity variants work poorly with some types of 5' modified sgRNAs[103,104]. This issue has technical aspects: to comply with the sequence requirement of the promoters commonly used to transcribe the sgRNA (the human U6 promoter in mammalian cells[36,44,108-110] or the T7 promoter *in vitro*[111-113]) 5' modified sgRNAs are frequently used with the WT SpCas9 when appropriate 20G-N19-NGG targets cannot be identified bioinformatically. If sgRNAs are delivered as plasmids, the most commonly used promoter for the transcription - if not tissue-specific expression is needed - is the RNA polymerase III driven human U6 promoter[36,44,114,115]. The U6 promoter requires a G nucleotide at the +1 position for proper transcription, although it has been suggested that the promoter works properly with an initial adenine as well, but the thorough characterization of the requirements of U6 promoter is still missing[116,117]. The other commonly used promoter if *in vitro* transcribed sgRNAs are needed (for example for RNP complex formation) is the T7 promoter, which is driven by T7 RNA polymerase. It seems that there is no consensus if only one G nucleotide or GG

nucleotides are required for the T7 polymerase for the preferred starting sequences of the transcript[111-113].

Another observation is that increased fidelity variants have a lower on-target activity on some target sites compared to WT SpCas9[103,104,106,107]. Furthermore, it has also been reported that increased fidelity variants do not work effectively in the pre-assembled RNP form that would be relevant to their clinical application[118-121]. In summary, despite these limitations, increased fidelity SpCas9 variants clearly are one of the best approaches for precise genome editing and their further developments are particularly important.

2. Aim of this study

While the results seen with increased fidelity variants are very encouraging, it has become clear that there are still several target sequences which can be cleaved by increased fidelity variants only with considerable off-target effect. Unfortunately, it is hardly predictable which sequences these are.

Furthermore, it is difficult to decide which SpCas9 variant is superior for applications where the avoidance of off-target activity is of paramount interest, because they were characterized in differing experimental setups: exploiting different sets of targets in different cells and employing different methods to assess their genome-wide specificity.

Even though the propensity for off-target activity has been considerably decreased of SpCas9 in general by using its increased fidelity variants, there are several limitations detailed in the introduction (e.g. their lower activity with 5' modified sgRNAs or when used in RNP form, as well as their reduced target space) which prevent the exploitation of their full potential.

Therefore, I pursued the following aims:

1. Systematically characterizing the increased fidelity variants in the same experimental sets and conditions.
2. Making a new, even better working increased fidelity SpCas9 variant with even higher fidelity than the already published ones.
3. Examining the compatibility between increased fidelity variants and different type of 5' modified sgRNAs.
4. Making new increased fidelity variants which can be used with 5' modified sgRNAs.
5. Examining how the specificity of genome modification using the increased fidelity variants could be further increased by combining it with other fidelity increasing approaches.

3. Materials and methods

3.1. Materials

Restriction enzymes, T4 ligase, Dulbecco's modified Eagle Medium DMEM (Gibco), fetal bovine serum (Gibco), Turbofect, Lipofectamine 2000, TranscriptAid T7 High Yield Transcription Kit, Qubit™ dsDNA HS Assay Kit, Shrimp Alkaline Phosphatase (SAP), Taq DNA polymerase (recombinant), Platinum Taq DNA polymerase and penicillin/streptomycin were purchased from Thermo Fischer Scientific, protease inhibitor cocktail was purchased from Roche Diagnostics. DNA oligonucleotides, trimethoprim (TMP) and GenElute HP Plasmid Miniprep kit were acquired from Sigma-Aldrich. ZymoPure Plasmid Midiprep kit and RNA Clean & Concentrator kit were purchased from Zymo Research. NEBuilder HiFi DNA Assembly Master Mix and Q5 High-Fidelity DNA Polymerase were obtained from New England Biolabs Inc. NucleoSpin Gel and PCR Clean-up kit was purchased from Macherey-Nagel. 2 mm electroporation cuvettes were acquired from Cell Projects Ltd, Bioruptor 0.5 ml Microtubes for DNA Shearing from Diagenode. Agencourt AMPure XP beads were purchased from Beckman Coulter. T4 DNA ligase (for GUIDE-seq) and end-repair mix were acquired from Enzymatics. KAPA universal qPCR Master Mix was purchased from KAPA Biosystems.

3.2. Plasmid construction

Vectors were constructed using standard molecular biology techniques including the one-pot cloning method[122], E. coli DH5α-mediated DNA assembly method[123], NEBuilder HiFi DNA Assembly and Body Double cloning method[124]. For detailed sequence information about sgRNA target sites and mismatching sgRNAs sequences are available in Appendix: App. Table 1, 2 and 3. The sequences of all plasmid constructs were confirmed by Sanger sequencing (Microsynth AG). Plasmids developed by us are listed in App. Table 4 and 5 and plasmids acquired from the non-profit plasmid distribution service Addgene (http://www.addgene.org/) are listed in App. Table 5.

3.3. Cell culturing and transfection

Cells employed in the studies are N2a (neuro-2a mouse neuroblastoma cells, ATCC – CCL-131), HEK293 (Gibco 293-H cells), N2a.dd-EGFP (cell line developed by us containing a single integrated copy of an EGFP-DHFR[DD] [EGFP-folA dihydrofolate reductase destabilization domain][125] fusion protein coding cassette driven by the Prnp promoter) as well as N2a.EGFP and HEK-293.EGFP (both cell lines containing a single integrated copy of an EGFP cassette driven by the Prnp promoter) cells. Cells were grown at 37 °C in a humidified atmosphere of 5% CO_2 in high glucose Dulbecco's Modified Eagle medium (DMEM) supplemented with 10% heat inactivated fetal bovine serum, 4 mM L-glutamine (Gibco), 100 units/ml penicillin and 100 µg/ml streptomycin. Cells were plated one day prior to transfection in 48-well plates at a density of approximately 25,000-30,000 cells/well unless otherwise noted. Transfections were performed with TurboFect transfection reagent according to the manufacturer's recommended protocol.

3.4. Flow cytometry

Flow cytometry analyses were carried out on Attune Acoustic Focusing Cytometer (Applied Biosystems), Attune NxT Acoustic Focusing Cytometer (Applied Biosystems) or CytoFLEX Flow Cytometer (Beckman Coulter). For data analysis Attune Cytometric Software v.2.1.0, Attune NxT Software v.2.7.0 or CytExpert 2.0 were used, respectively. Viable single cells were gated based on side and forward light-scatter parameters and a total of 5,000 - 10,000 viable single cell events were acquired in all experiments. The following instrument parameters were used: Attune Acoustic Focusing Cytometer parameters: the EGFP fluorescence signal was detected using the 488 nm diode laser for excitation and the 530/30 nm filter for emission, the mCherry fluorescent signal was detected using the 488 nm diode laser for excitation and a 640LP filter for emission. Attune NxT Focusing Cytometer parameters: the EGFP fluorescence signal was detected using the 488 nm diode laser for excitation and the 530/30 nm filter for emission, the mCherry fluorescent signal was detected using the 561 nm diode laser for excitation and a 620/15 nm filter for emission. CytoFLEX Flow Cytometer parameters: the EGFP fluorescence signal was detected using the 488 nm diode laser for excitation and the 525/40 nm filter for emission, the mCherry fluorescent signal was detected using the 638 nm diode laser for excitation and a 660/20 filter for emission.

3.5. EGFP disruption assay

In EGFP disruption assay we are detecting the loss of fluorescence signal in cells caused by frameshift mutations from the error-prone non-homologous end joining (NHEJ)-mediated repair after the targeted double-stranded breaks in a single copy integrated EGFP reporter gene (App. Fig. 1). All EGFP disruption experiments were conducted in N2a.dd-EGFP cells except the on-target screen (Fig. 4A) and results shown in Figure 3, 5-7 and App. Figure 4A, which was conducted on N2a.EGFP cells (see details below). Transfection efficacy was calculated via mCherry expressing cells. Transfections were performed in triplicate.

In case of the **On-target screen (Fig. 4A, App. Fig. 4A)** N2a.EGFP cells were co-transfected with two types of plasmids: SpCas9 variant expression plasmid (137 ng) and sgRNA and mCherry coding plasmid (97 ng) using 1 µl TurboFect reagent per well in 48-well plates. Transfected cells were analyzed 4 days post-transfection by flow cytometry. In this cell line the EGFP disruption level was not saturated, this way this assay is a more sensitive reporter of the intrinsic activities of these nucleases compared to N2a.dd-EGFP cell line.

In the case of a **Mismatch screen (such as Fig. 4B)** N2a.dd-EGFP cells were co-transfected with two types of plasmids: with SpCas9 variant expression plasmid (137 ng) and a mix of 3 sgRNAs in which one nucleotide position was mismatched to the target using all 3 possible bases and mCherry coding plasmid (3x∼33.3 ng = 97 ng) using 1 µl TurboFect reagent per well in 48-well plates. TMP (trimethoprim; 1 µM final concentration) was added to the media ∼48 h before FACS analysis. Transfected cells were analyzed 4 days post-transfection by flow cytometry. The 4 days post-transfection in this cell line shows a close to saturated level, this way it is a good reporter system to see the full spectrum of off-target activities. Background EGFP loss for each experiment was determined using co-transfection of dead SpCas9 (dSpCas9) expression plasmid and different targeting sgRNA/mCherry coding plasmids.

In case of the experiments conducted on **N2a.EGFP** (Fig. 3, 5-7 and App. Fig. 1) cells were co-transfected with two types of plasmids: SpCas9 expression plasmid (137 ng) and sgRNA and mCherry coding plasmid (78-97 ng) using 1 µl TurboFect reagent per well in 48-well plates. Transfected cells were analysed 3 and 7 days post-transfection

by flow cytometry. Transfection efficacy was calculated via mCherry expressing cells measured 3 days post-transfection. EGFP positive cells were counted 7 days post-transfection. Background level of EGFP for each experiment was determined by using co-transfection of a dSpCas9 expression plasmid and a targeted sgRNA/mCherry coding plasmid. The 7 days post-transfection in this cell line showed a close to saturated level, this way it is a good reporter system to see the highest on-target activity level reachable by the sgRNA-SpCas9 variant complex.

EGFP disruption values were calculated as follows: the average EGFP background loss from dSpCas9 control transfections made in the same experiment was subtracted from each individual treatment in that experiment and the mean values and standard deviation (s.d.) were calculated from it. In the case of normalization, the results were normalized to the WT SpCas9 data from the same experiment.

3.6. Western blot

N2a.dd-EGFP cells were cultured in 48-well plate and were transfected as described above in the EGFP disruption assay section. Four days post-transfection, 9 parallel samples corresponding to each type of SpCas9 variant transfected were washed with PBS, then trypsinized and mixed, and were analyzed for transfection efficiency via mCherry fluorescence level by using flow cytometry. The cells from the mixtures were pelleted (200 g, 5 min at 4 °C). Pellets were resuspended in ice cold Harlow buffer (50 mM Hepes pH 7.5; 0.2 mM EDTA; 10 mM NaF; 0.5% NP40; 250 mM NaCl; Protease Inhibitor Cocktail 1:100; Calpain inhibitor 1:100; 1 mM DTT) and lysed for 20-30 min on ice. The cell lysates were centrifuged at 19,000 g for 10 min. The supernatants were transferred into new tubes and total protein concentrations were measured by the Bradford protein assay. Before SDS gel loading, samples were boiled in Protein Loading Dye for 10 min at 95 °C. Proteins were separated by SDS-PAGE using 7.5% polyacrylamide gels and were transferred to a PVDF membrane using a wet blotting system (Bio-Rad). Membranes were blocked by 5% non-fat milk in Tris buffered saline with Tween20 (TBST) (blocking buffer) for 2 h. Blots were incubated with primary antibodies [anti-FLAG (F1804, Sigma) at 1:1,000 dilution; anti-β-actin (A1978, Sigma) at 1:4,000 dilution in blocking buffer] overnight at 4 °C. The next day after washing steps in TBST the membranes were incubated for 1 h with HRP-conjugated secondary anti-mouse

antibody 1:20,000 (715-035-151, Jackson ImmunoResearch) in blocking buffer. The signal from detected proteins was visualized by ECL (Pierce ECL Western Blotting Substrate, Thermo Scientific) using a CCD camera (Bio-Rad ChemiDoc MP).

3.7. Transcriptional activation

One of the applications of SpCas9 nucleases, besides being a programmable nuclease, is using it for delivering effector domains precisely to a chosen locus within the genome for modulating endogenous gene expression[11,126-133]. For transcriptional activation we exploited a method published by Konermann and colleagues, that uses a modified sgRNA containing two minimal hairpin aptamer - which selectively binds dimerized MS2 bacteriophage coat proteins[134]- in the tetraloop and stem loop 2. The sgRNA complexed with dSpCas9 binds to the target site and the hairpins recruit the MS2-p65-HSF1 fused protein which can mediate transcriptional upregulation.[133] For applying this method in our experiments, N2a.EGFP cells were co-transfected with three types of plasmids as follows: 91 ng of SpCas9 expressing plasmid, 83 ng of the mixture of five sgRNAs (targeting the *Prnp* promoter sequence [for details see App. Table 2]) coding plasmid which expresses MS2-p65-HFS1 fusion protein as well and 75 ng of mCherry coding plasmid (pcDNA3-mCherry) using 1 μl TurboFect reagent per well in 48-well plates. Transfected cells were analysed 3 days post-transfection. The transfection efficacy was calculated via mCherry fluorescence level. The relative upregulation was calculated from the median of the EGFP intensity. Background level of EGFP was determined using a negative control for transfections: a mock plasmid ([MS2-p65-HFS1_sgRNA(MS2)] without spacer sequence) co-transfected with SpCas9 coding plasmid and with mCherry coding plasmid. The medians obtained for EGFP fluorescence were averaged between three parallel samples and the error was estimated by Gaussian error propagation of the component errors (s.d.) associated to the measured variables.

3.8. *In vitro* transcription

Most of the *in vitro* transcriptions were performed by Zoltán Ligeti. sgRNAs were *in vitro* transcribed using TranscriptAid T7 High Yield Transcription Kit and PCR-generated double-stranded DNA templates carrying a T7 promoter sequence. Primers used for the preparation of the DNA templates are listed in App. Table 6. sgRNAs were

dephosphorylated with SAP, purified with the RNA Clean & Concentrator kit, and reannealed (95 °C for 5 min, ramp to 4 °C at 0.3 °C/sec). sgRNAs were quality checked using 10% denaturing polyacrylamide gels and ethidium bromide staining.

3.9. Protein purification

The protein expression and purification was performed by Zoltán Ligeti. All active SpCas9 variants were subcloned into the pMJ806 plasmid, the dead variants were cloned into the pET-dCas9-VP64-6xHis plasmid. SpCas9 variants subcloned by us are listed in App. Table 4 and 5.

The expression constructs of the SpCas9 variants were transformed into BL21 Rosetta 2 (DE3) cells, grown in Luria-Bertani (LB) medium at 37 °C for 16 h. 10 ml from this culture was inoculated into 1 l of growth media (12 g/l Tripton, 24 g/l Yeast, 10 g/l NaCl, 883 mg/l NaH2PO4 H2O, 4.77 g/l Na2HPO4, pH 7.5) and cells were grown at 37 °C to a final cell density of 0.6 OD600, and then were chilled at 18 °C. The protein was expressed at 18 °C for 16 h following induction with 0.2 mM IPTG. The protein was purified by a combination of chromatographic steps by NGC Scout Medium-Pressure Chromatography Systems (Bio-Rad). The bacterial cells were centrifuged at 6,000 rcf for 15 min at 4 °C. The cells were resuspended in 30 ml of Lysis Buffer (40 mM Tris pH 8.0, 500 mM NaCl, 20 mM imidazole, 1 mM TCEP) supplemented with Protease Inhibitor Cocktail (1 tablet/30 ml; complete, EDTA-free, Roche) and sonicated on ice. Lysate was cleared by centrifugation at 48,000 rcf for 40 min at 4 °C. Clarified lysate was bound to a 5 ml Mini Nuvia IMAC Ni-Charged column (Bio-Rad). The resin was washed extensively with a solution of 40 mM Tris pH 8.0, 500 mM NaCl, 20 mM imidazole, and the bound protein was eluted by a solution of 40 mM Tris pH 8.0, 250 mM imidazole, 150 mM NaCl, 1 mM TCEP. 10% glycerol was added to the eluted sample.

In case of the inactive SpCas9 variants the protein was dialyzed 2x1 h against 20 mM HEPES pH 7.5, 150 mM KCl, 1 mM DTT, 1% glycerol. The dialyzed protein was purified on a 3x1 ml Bio-Scale Mini Macro-Prep High S column (Bio-Rad), eluting with 1 M KCl, 20 mM HEPES pH 7.5, 1 mM DTT.

In case of the active SpCas9 variants the 6xHis-MBP-TEV fusion protein was cleaved by TEV protease (3 h at 25 °C). The volume of the protein solution was made up to 100 ml with buffer (20 mM HEPES pH 7.5, 100 mM KCl, 1 mM DTT). The cleaved

protein was purified on a 5 ml HiTrap SP HP cation exchange column (GE Healthcare) and eluted with 1 M KCl, 20 mM HEPES pH 7.5, 1 mM DTT. The active and dead SpCas9 proteins were further purified by size exclusion chromatography on a Superdex 200 10/300 GL column (GE Healthcare) in 20 mM HEPES pH 7.5, 200 mM KCl, 1 mM DTT and 10% glycerol. The eluted protein was confirmed by SDS-PAGE and Coomassie brilliant blue R-250 staining. The protein was stored at -20 °C.

3.10. Electrophoretic mobility shift assay (EMSA)

Electrophoretic mobility shift assay is a commonly used *in vitro* technique to study protein-DNA interactions. Using electrophoretic separation one can decide if binding occurs between the protein and the DNA sequence by shifting the apperence of the DNA band on the gel. The EMSA experiments were performed by Zoltán Ligeti. DNA target sequences were amplified from an EGFP-KDEL coding sequence template (for detailed primer and sequence information see App. Table 7). Binding assays were performed in buffer containing 20 mM HEPES pH 7.5, 100 mM KCl, 5 mM $MgCl_2$, 0.1 mg/ml heparin, and 1 mM TCEP in a total volume of 20 µl. The sgRNA was supplied two-fold the molar amount of protein. The target DNA (40 nM) was incubated with protein-sgRNA complex (160 nM) (or in case of the serial dilution experiments with 160, 320, 640 and 1280 nM protein-sgRNA complex concentrations, respectively) for 30 min at 37 °C. Samples were resolved at 4 °C on an 8% native polyacrylamide gel containing 0.5X TBE and 10 mM $MgCl_2$. The gel was stained with ethidium bromide.

3.11. EGFP disruption assay with RNP

The EGFP disruption assay with RNP experiments were performed by András Tálas. N2a.dd-EGFP cells cultured in 48-well plates, were seeded a day before transfection at a density of 3×10^4 cells/well in 250 µl complete DMEM. 13.75 pmol SpCas9 and 16.5 pmol sgRNA was complexed in Cas9 storage buffer (20 mM HEPES pH 7.5, 200 mM KCl, 1 mM DTT and 10 % glycerol) for 15 minutes at RT. 25 µl serum-free DMEM and 0.8 µl Lipofectamine 2000 was added to the complexed RNP and incubated for 20 minutes prior to adding to the cells. TMP (trimethoprim; 1 µM final concentration) was added to the media ~48 h before FACS analysis. Transfected cells were analyzed 4 days post-transfection by flow cytometry. Transfections were performed

in triplicate. Background EGFP loss for each experiment was determined using co-transfection of WT SpCas9 expression plasmid and non-targeted sgRNA and mCherry coding plasmids. EGFP disruption values were calculated as follows: the average EGFP background loss from control transfections made in the same experiment was subtracted from each individual treatment in that experiment and the mean values and standard deviation (s.d.) were calculated from it.

3.12. GUIDE-seq

Genome-wide, Unbiased Identification of DSBs Enabled by sequencing (GUIDE-seq) is a method which allows the unbiased identification of the off-target sites occurred after a targeted DSB (such as the usage of SpCas9). The method relies on the efficient integration of an end-protected, short, double-stranded oligodeoxynucleotide (dsODN) into sites of DSBs in living cells. The genomic context around each off-target site is amplified using this short sequence (and the ligated adapters to the DNA ends after the *in vitro* shearing), followed by NGS and *in silico* identification of the occurring off-target cleavage sites.[75] We performed GUIDE-seq experiments with different SpCas9 variants on 13 different target sites. Briefly, 2×10^6 HEK293.EGFP cells were transfected either (1) with 3 µg of SpCas9 variant expressing plasmid, 1.5 µg of mCherry and sgRNA coding plasmid or (2) with 100 pmol SpCas9 and 120 pmol sgRNA, which was complexed in Cas9 storage buffer (20 mM HEPES pH 7.5, 200 mM KCl, 1 mM DTT and 10% glycerol) for 15 minutes at RT. 100 pmol of the dsODN containing phosphorothioate bonds at both ends (according to the original GUIDE-seq protocol[75]) was mixed together with 100 µl home-made nucleofection solution to the plasmid or RNP complex and electroporated as described in Vriend et al.[135] using Nucleofector (Lonza) with A23 program and 2 mm electroporation cuvettes. Transfected cells were analyzed 3 days post-transfection by flow cytometry. Genomic DNA was purified according to Puregene DNA Purification protocol (Gentra systems). Genomic DNA was sheared with BioraptorPlus (Diagenode) to 550 bp in average. Sample libraries were assembled as previously described[75] and sequenced on an Illumina MiSeq instrument by ATGandCo Ltd. Data were analyzed using open-source guideseq software (version 1.1)[136]. Consolidated reads were mapped to the human reference genome GrCh37 supplemented with the integrated EGFP sequence. Upon identification of the genomic regions integrating double-stranded

oligodeoxynucleotide (dsODNs) in aligned data, off-target sites were retained if at most seven mismatches against the target were present and if absent in the background controls. Visualization of aligned off-target sites are provided as a color-coded sequence grid.

3.13. TIDE

Tracking of Indels by DEcomposition (TIDE) method [81] was applied for analyzing mutations and determining their frequency in a cell population using different sgRNAs and SpCas9 proteins. From the isolated genomic DNA PCR was conducted with Q5 High-Fidelity DNA Polymerase in triplicates (for PCR primer details, see App. Table 8). Genomic PCR products were gel excised via NucleoSpin Gel and PCR Clean-up kit and were Sanger sequenced. Indel efficiencies were analysed by TIDE webtool (https://tide.nki.nl/) by comparing SpCas9 treated and control samples.

3.14. HR mediated integration assay

We used a donor plasmid containing an EGFP cassette flanked by 1,000 bp-long homology arms identical to the target gene to integrate it precisely into the targeted site by HDR. We performed the experiment as follows: N2a cells were seeded into 48-well plates a day before transfection at a density of 2.5×10^4 cells/well. Next day cells were co-transfected with three types of plasmids: an expression plasmid for EGFP flanked by 1,000 bp-long homology arms to the *Sprn* gene (166 ng), SpCas9 variant expressing plasmid (42 ng) and an sgRNA/mCherry coding plasmid (42 ng), giving 250 ng total plasmid DNA, using 1 µl TurboFect reagent per well. Transfected cells were analyzed 4 and 18 days post-transfection by flow cytometry. Transfection efficiency was calculated via mCherry expressing cells measured 4 days post-transfection. EGFP positive cells were counted 18 days post-transfection. Transfections were performed in triplicate.

3.15. NHEJ-mediated integration using a ‘self-cleaving’ EGFP-expression plasmid

The key feature of this approach is the use of a ‘self-cleaving’ plasmid that enhances targeted integration[137]. In case of this method we are transfecting two plasmids: (i) a plasmid expressing the SpCas9 and an sgRNA targeting the genomic site and (ii) a ‘self-cleaving’ donor plasmid harboring the EGFP cassette and a - genomic-targetless – sgRNA targeting the donor plasmid itself. After the transfection the SpCas9 does not only

cleave the genomic target site but cleaves the donor plasmid as well. During the repair of the broken genomic DNA, the linear donor plasmid will more efficiently integrate into the genome than a circular donor plasmid.

We performed the experiment as follows: N2a cells were seeded into 12-well plates a day before transfection at a density of 8×10^4 cells/well. Next day cells were co-transfected with three types of plasmids: a 'self-cleaving' promoterless-EGFP-expression plasmid [137] (which has to integrate in-frame for *Sprn* promoter driven EGFP expression) (1 μg), SpCas9 expressing plasmid (590 ng) and an sgRNA/mCherry coding plasmid (410 ng), giving 2 μg total plasmid DNA, using 4 μl TurboFect reagent per well. Transfections were performed in triplicate. Transfection efficiency was calculated via mCherry expressing cells measured 4-days post-transfection. EGFP positive cells were counted 14-days post-transfection. The 'self-cleaving' promoterless-EGFP-expression plasmid also contained an EF1-alpha promoter driven mCherry expressing cassette to monitor the integration of the linearized plasmid 14-days post-transfection.

3.16. Statistics

Differences between SpCas9 variants were tested by using either Paired-samples Student's t-test (Fig. 11A: evoSpCas9/B-evoSpCas9; Fig. 11B: WT SpCas9 / B-WT SpCas9, eSpCas9 / B-eSpCas9; Fig. 11C: evoSpCas9/B-evoSpCas9, HeFSpCas9/B-HeFSpCas9) or by using Wilcoxon Signed Ranks test (in case of all Blackjack- parental variant pairs in Figure 11A-C not mentioned above) in cases where differences did not meet the assumptions of Paired t-test. Normality of data and of differences were tested by Shapiro-Wilk normality test. Statistical tests were performed using IBM SPSS ver. 20 on data including all parallel sample points.

Pearson correlation was calculated by Elfrieda Fodor. Person correlation matrix analysis with significance was done by using R language and environment, version 3.4.1 (C: 2017 The R Foundation for Statistical Computing) and the package corrplot, version. 0.84 (2017).

4. Results

As the first step of this study, we aimed to characterize and compare the wild type (WT), the already existing increased fidelity SpCas9 variants (e-, -HF1, Hypa- and evoSpCas9) and HeFSpCas9 what we made by combining the mutations of eSpCas9 and SpCas9-HF1 in the same experimental set ups. The idea behind the combination of the mutations in HeFSpCas9 were that both rationally designed SpCas9 variants (e- and SpCas9-HF1) were made based on different working rationales, namely the interactions of the protein with different strands of the target DNA are altered by the mutations. Therefore, we hypothesized that the mutations could have synergetic effects, and that HeFSpCas9 could have even higher fidelity than their parental variants. In order to facilitate a thorough comparison of the nuclease variants, we used the same codon optimization for the wild type and for all mutant SpCas9 variant and tailored them to have identical NLS and FLAG tags at their termini in the same mammalian expression plasmid backbone (Fig. 2A). After the molecular cloning of the plasmids, we tested them by western blotting which indicated that the mutations do not alter the expression level of SpCas9 and the amounts of the proteins expressed at steady state are comparable (Fig. 2B).

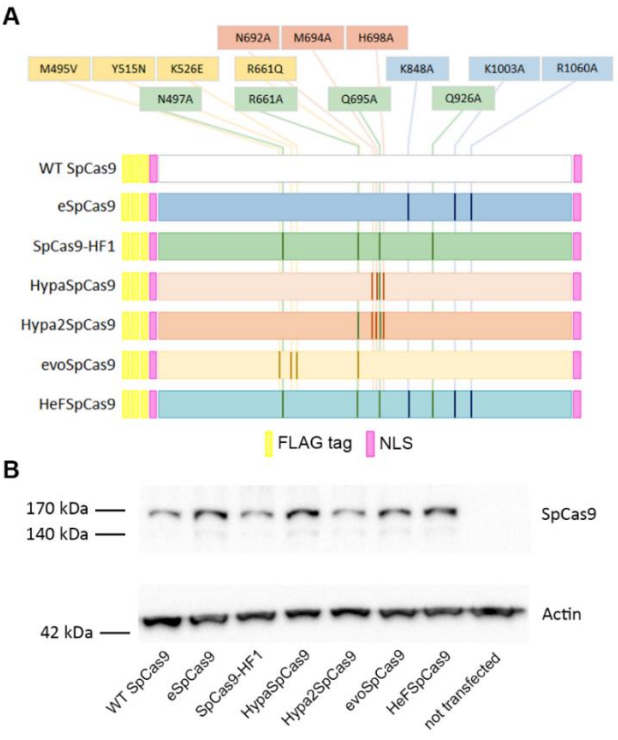


Figure 2. SpCas9 variants employed in these studies and their expression levels.

A, Schematics depicting the main features of the wild type and the SpCas9 variants used: each protein sequence is flanked by a nuclear localization signal (NLS) at both ends and is preceded by a 3xFLAG tag. For detailed information about the amino acid mutations in the different SpCas9 variants see App. Table 4. **B,** Immunoblot analysis of the expression levels of SpCas9 nuclease variants (~160 kDa) in cell lysates of reporter N2a.dd-EGFP cells transfected with the indicated nuclease constructs. β-actin (~42 kDa) was used as a control for total protein amounts analyzed.

4.1. <u>There is a ranking by fidelity and target selectivity among the increased fidelity SpCas9 variants and a ranking among the DNA targets.</u>

We decided that we will perform our first direct comparisons of the SpCas9 variants using an EGFP disruption assay (App. Fig. 1). Firstly, we tested the on-target activity of the SpCas9 variants by employing 47 sgRNAs (for detailed sequence information see App. Table 1) with perfectly matching 20 nucleotide-long spacers targeting the EGFP coding sequence. Secondly, we characterized the mismatch tolerance of the increased fidelity variants by comparing their activity on target sites with single base mismatches, because it is a sensitive approach capable of discerning small fidelity differences among the increased fidelity variants. We placed the mismatches in the PAM distal regions (between the 19th to 14th positions) of the spacer sequence, where mismatches are most tolerated by SpCas9. We investigated 16 EGFP target sites (out of the 47 tested in the on-target screen) by employing 144 mismatching sgRNAs: each target with a matching and with 9 one-base mismatching sgRNAs (three different positions in case of each spacer and all three possible mismatches at each position, as shown in Figure 3A; for detailed sequence information see App. Table 3). Since we found that mixing the three possible sgRNAs mismatching the same position resulted in a sensitive reporting of the off-target activities in the EGFP disruption assay, we used this approach here (Fig. 3B).

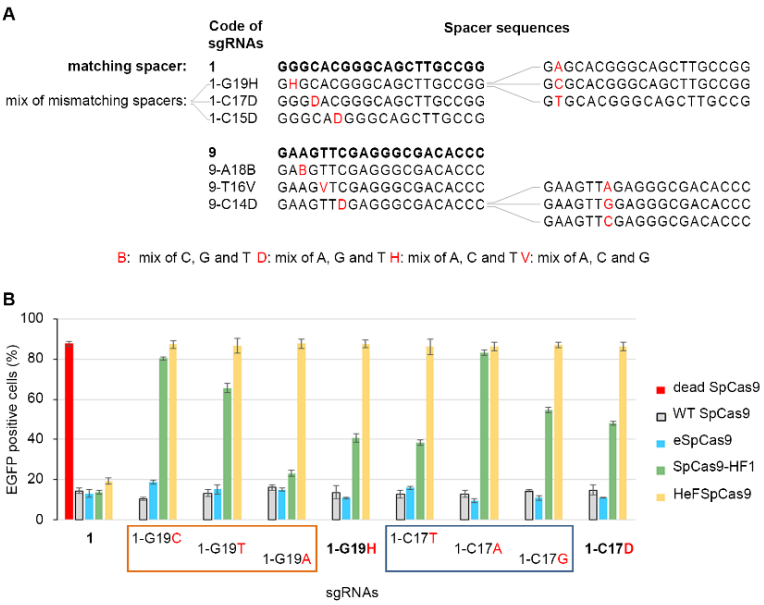


Figure 3. Mixing of the sgRNAs mismatching the same position. A, Examples of the mismatching sgRNAs; in case of each target site with a matching and with 9 one-base mismatching sgRNAs: each of the three possible mismatches for three different positions. **B,** EGFP disruption experiments with sgRNAs mismatched at identical positions loaded to SpCas9 nucleases individually or as a mixture of the three possible mismatching variants. Lower percentage of EGFP positive cells means higher level of disruption,

namely higher level of nuclease activity. Mixing the sgRNAs mismatching at the same position resulted in a sensitive reporting closely with the average of the outcome from the three individually loaded sgRNAs' off-target activities. Means are shown, error bars represent the standard deviation (s.d.) for n=3 biologically independent samples.

The results of the on-target and mismatch screen experiments revealed a very intriguing pattern (Fig. 4A and B) with important consequences. The fidelity of the variants gradually increases in parallel with increasing target-selectivity (i.e. the variant does not cleave or cleaves with reduced activity the target sites that are cleaved by the WT SpCas9 and that results in decreasing average activity and shrinking target space) in the following order: WT SpCas9 < eSpCas9 < SpCas9-HF1 < HypaSpCas9 < evoSpCas9 < HeFSpCa9. Interestingly, these differences among the variants were not clear in the original publications.

Whereas wild type SpCas9 barely distinguishes perfect matches from single mismatches on the targets examined, the other SpCas9 variants are capable of strong discrimination in a target dependent manner. Among the tested nucleases, the fidelity/activity of the first three increased fidelity variants (e-, -HF1 and HypaSpCas9) clustered most closely to each other. Since the greatest difference in fidelity was between HypaSpCas9 and evoSpCas9, we decided to create an increased fidelity variant optimized for targets in the in-between range and to this end we constructed Hypa-R661A-SpCas9 (referred to as Hypa2SpCas9) by introducing the R661A mutation into its sequence, a mutation common to both SpCas9-HF1 and evoSpCas9 (Fig. 2A). This variant was also included in the screens shown in Figure 4A and 4B. The results seen with HeFSpCas9 were confirmed in a recent publication as well[106].

Remarkably, the cleavability of different targets by the various increased fidelity variants is also strongly related to the fidelity of the variants, allowing the target sequences to be ranked as well (Fig. 4A). This target ranking is also apparent in the mismatch screen (Fig. 4B) indicating that the efficiency and the specificity of target-cleavage by the increased fidelity SpCas9 variants are interrelated and highly determined by the targets. We suggest that this pattern seen on the heatmaps is caused by the cumulative effect of (i) the contribution of the specific target, (ii) the increased-fidelity mutations and (iii) the mismatches, if there are any (Fig. 4C). The combinations of these

effects regulate the cleavage activity in each specific cleavage event and explains the results seen in Figure 4.

Taking a closer look at the target ranking it can be seen that at one end of the ranking there are sequences (we named them "*high-contributing"* target sites) that all variants can cleave, but only the highest fidelity variant cleaves them with minimal off-target cleavage (such as EGFP target sites 1 and 43 in Fig. 4A-C; the other variants cleave them with high off-target effects). Interestingly, although the highest fidelity nucleases show decreased overall on-target activity (increased target-selectivity), they are capable of undiminished activity, comparable to that of WT SpCas9 on a small subset of targets, particularly, on those targets, which are exactly the ones falling into the *high-contributing* target range. At the other end of the ranking there are sequences ("*low-contributing*") that are cleaved efficiently only by the lower increased fidelity variants and by the WT SpCas9 (such as EGFP target sites 6, 8 and 24 in Fig. 4A-C) but also with significantly lower level of off-target activity. The higher fidelity variants, such as HeF-, evo-, Hypa2SpCas9 do not cleave these *low-contributing* target sites effectively.

This propensity of the targets makes efforts to develop a superior nuclease variant that cleaves all targets with high on-target activity and high specificity (without any off-target effect) futile. Instead, our results show that it is necessary to have a series of variants with increasing fidelity in order to have an optimal increased fidelity nuclease for targets in all cleavability ranks (from *low*-to-*high contributing* target sites; Fig. 4C).

To confirm that these patterns are not specific only for the N2a.dd-EGFP mouse cell line, we selected 9 targets of various ranks and tested them using four SpCas9 variants (WT, e-, -HF1 and HeFSpCas9) in a mismatch screen on HEK-293.EGFP human and on N2a.EGFP mouse cell lines. The patterns seen in case of both cell lines are almost identical to each other (data not shown here, Kulcsár et al. Additional file 1: Figure S4)[138] and are similar to those shown in Figure 4B, supporting the idea that the features of the increased fidelity nucleases and that of the targets, which have become apparent in this study, are their intrinsic characteristics and are not specific to the particular cell line used in the experiments.

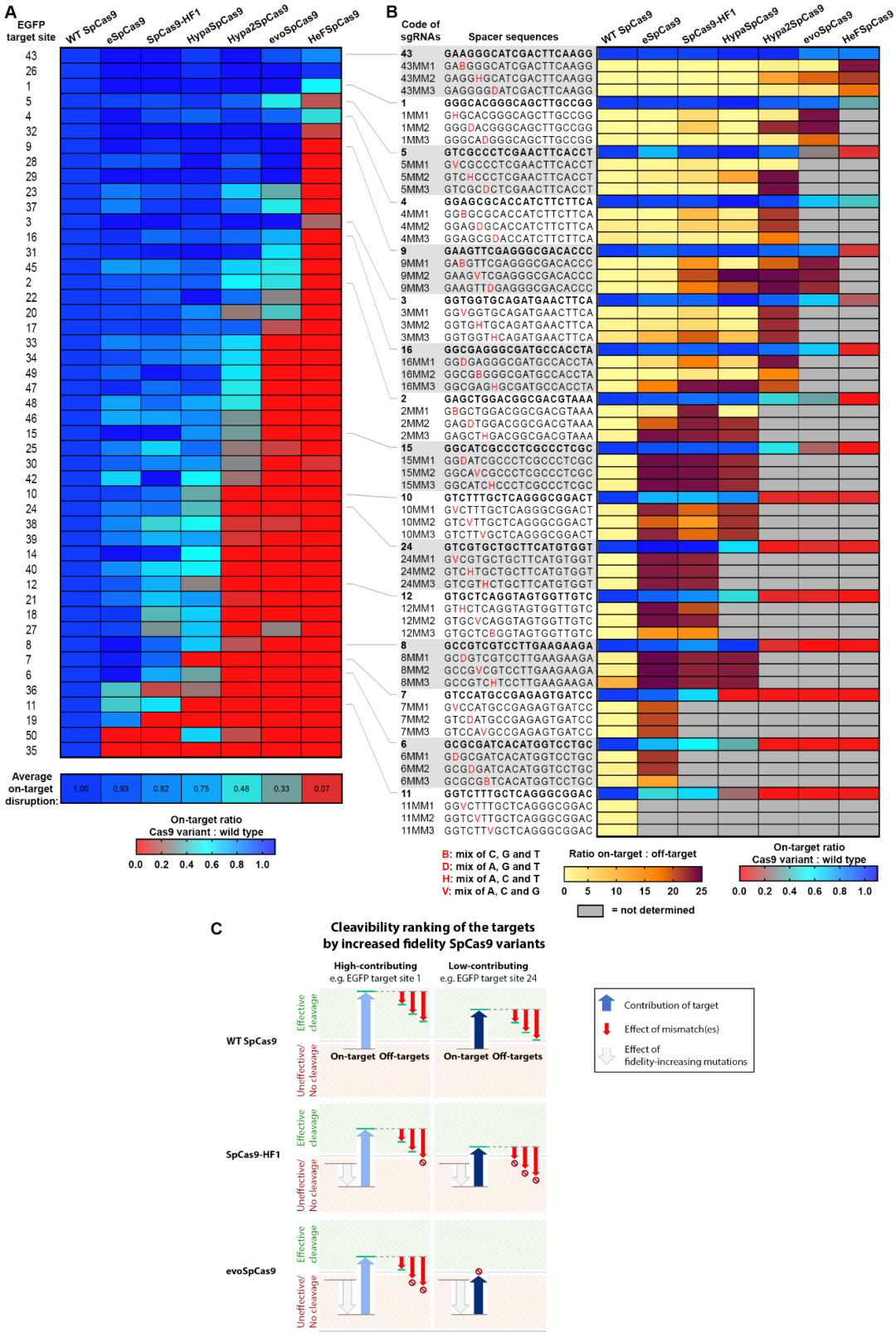


Figure 4. There is a target-selectivity and a fidelity ranking among the SpCas9 variants and a cleavability ranking among the target sites. EGFP disruption activities of SpCas9 nucleases. **A,** Heatmap showing the relative on-target activities (red to blue) of increased fidelity nucleases

normalized to that of the wild type for each of the targets in N2a.EGFP cells. The average on-target activities are shown under the heatmap. Some outliers have been confirmed in repeated experiments. They indicate that other factors besides increased fidelity protein mutations and target contributions modulate the on-target activities of increased fidelity nucleases as well. However, the effects of these two main factors, the *target-selectivity* ranking of the nuclease variants and the *cleavability* ranking of the targets are clearly visible from the pattern. **B,** Mismatch screen of the nuclease variants with sgRNAs with either perfectly matching 20G-sgRNA (red to blue) or partially mismatching 20G-sgRNAs (yellow to brown; a mixture of three different sgRNAs with a mismatch in the same position) as indicated in the figure, reveals the *fidelity* ranking of the variants. Grey boxes: not determined because of the on-target activity being too low. **C,** Schematics to explain the cleavage activities of the different increased fidelity SpCas9 variants on various targets seen in (A) and (B). The pattern on the heatmaps is taken shape because each target possesses a specific contribution (blue arrows) that combines with the effect of increased-fidelity mutations (grey arrows), and mismatches (red arrows) to regulate target cleavage.

4.2. 5' extended sgRNAs diminish the activities of increased fidelity nucleases more with a matching than with a mismatching G nucleotide.

5’ modified sgRNAs are frequently used with WT SpCas9 when appropriate 20G-N19-NGG targets cannot be identified bioinformatically in the desired genomic region, because the most commonly used promoters (such as the human U6 promoter in mammalian cells[36,44,108-110] or the T7 promoter for *in vitro* trancription[111-113]*)* require a starting 5’ G nucleotide for efficient transcription. Different practices exist for modifying the 5’-end of the spacer of the sgRNA to meet that starting G requirement such as (i) altering the 20th non-G 5'-end nucleotide to a G[139-141], (ii) extending the guide with an extra 5'-end 21st G nucleotide[44,54,142,143] (Fig. 1D) or (iii) using the spacers without alteration, i.e. with the 5'-end non-G nucleotide. In the latter case the RNA polymerase starts to transcribe efficiently the sgRNA when encounters the first suitable nucleotide. Thus, the majority of the sgRNAs in this case will not be 20 nt-long, with lengths depending on their 5'-end sequences[116,117]. (iv) Another approach to meet the requirements is to truncate back the spacer sequence until a G nucleotide is encountered (17-19 nucleotide long spacers still show reasonable on-target activity with WT SpCas9)[75,78,97,144]. We (data not shown here, Kulcsár et al. Additional file 1: Figure S1a,

b)[138] and others[103,104] found, that the increased fidelity nuclease variants performed poorly in experiments when sgRNAs with different 5' modified spacers were used.

Although there are some methods which circumvents the limitations of U6 and T7 promoters such as ribozyme-[145-148] or tRNA-flanked[115,149-151] sgRNAs or using chemically synthesized sgRNAs. These approaches are not general solutions to extend the target space available for increased fidelity SpCas9 variants beyond the 20G-N19-NGG target sequences. The ribozyme- and tRNA-fused sgRNA approaches have not been well characterized for the sequence dependence of sgRNA-processing. They have been shown to work in mammalian cells in only a few studies and their applicability has not been characterized thoroughly with valid comparisons. Thus, it is not clear whether, when an sgRNA in these studies works with reduced activity, it is due to the ribozyme or tRNA sequence addition or just due to the sequence dependence of the nucleases used. These uncertainties may be one of the reasons why these systems have not been applied to any large-scale studies and none of the pooled sgRNA libraries publicly available so far for knockout, activation or repression libraries are built on ribozyme- or tRNA-flanked sgRNA vectors. Chemically synthetized sgRNAs pre-assembled with the SpCas9 protein are a solution in some instances but are far from being a general solution. For example, RNPs are not suitable for use with pooled library screens, since the positive hits are identified through sequencing the incorporated lentivirus vectors and are prohibitively expensive for large-scale or high-throughput studies.

One can see that one of the potential limitations which could put a restraint on the broad applicability of the increased fidelity nucleases is the limited activity with 5' modified sgRNAs and so far there is no general alternative method available. Therefore, we decided that we will take a more thorough look on the most commonly used 5' modifications if they are compatible with increased fidelity variants.

We tested how the different 5' modified sgRNAs perform with two of the lower fidelity SpCas9 variants, namely eSpCas9 and SpCas9-HF1. In our experiments, we systematically examined how sgRNAs (i) with spacers mismatched with a 5' G at the 20th position, (ii) with 5' non-G spacers without alteration or (iii) with truncated spacers effect the cleavage activity of the increased fidelity variants in comparison to the WT protein. In case of all this three type of approaches the examined increased fidelity variants cleaved the on-target sites poorly, however, to different extents: eSpCas9 showed lower

sensitivity compared to SpCas9-HF1 (data not shown here, Kulcsár et al. Additional file 1: Figure S2e, f)[138]. These results are in line with earlier observations[103,104] and with papers published after our results were published[107,121,146,151].

More significantly, we also examined the perhaps most commonly used approach[44,54,142,143] to fulfill the G requirement: extending the sgRNA with an extra 21st 5' G nucleotide (21G-sgRNA, Fig. 1D). The effect of the 5' G extension of the spacer sequence was tested separately in case when the 21st G was (i) a non-matching G (Fig. 5A) or (ii) the G nucleotide was matching to the corresponding nucleotide in the targeted DNA strand (Fig. 5B). We chose 36 (26 for mismatching and 10 for matching 21st G nucleotide) EGFP target sites which starts with a perfectly matching 5' G at the 20th position to facilitate specifically investigating the effect under scrutiny. We compared the on-target activities of eSpCas9 and SpCas9-HF1 nucleases used with 21G-sgRNAs to the corresponding unmodified guide RNAs (20G-sgRNAs). The results of these experiments (Fig. 5) demonstrate that 21G-sgRNAs interfere with the activities of the increased fidelity nucleases. Our results were later confirmed by different studies[107,121,146,151]. Interestingly, we found that extending the guide with a matching 5' G nucleotide is much more detrimental to the activities of these nuclease variants than extending it with a mismatching one. In case of HypaSpCas9 we can see the same effect, namely matching 21G-sgRNAs are more detrimental than mismatching ones to the activities, but interestingly in case of higher fidelity variants (i.e. Hypa2-, evo- and HeFSpCas9) any kind of 5' extension is enough to completely spoil the cleavage activity (data not shown).

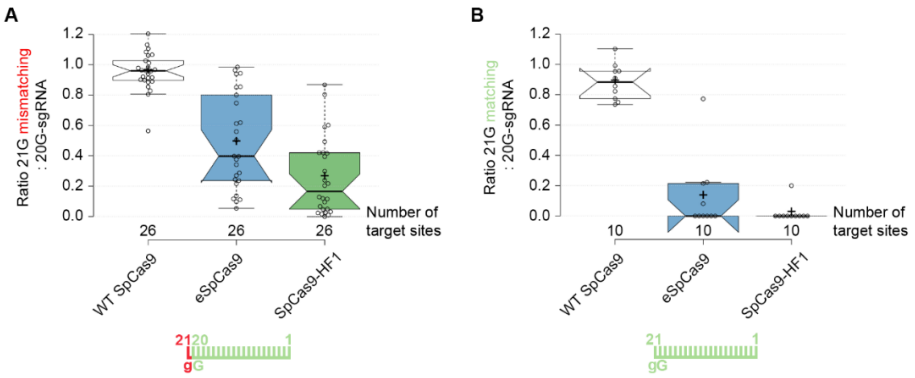


Figure 5. Extending the guide RNA with a matching 5' G nucleotide is much more detrimental to the activities than extending with a mismatching one in case of eSpCas9 and SpCas9-HF1. The experiment was partially carried out by Krisztina Huszár. Effect of 5' extension of the sgRNA with **A**, a mismatching G or **B**, a matching G nucleotide on the activities

of SpCas9 nucleases in comparison with using perfectly matching 20 nt-long spacers (20G-sgRNAs). Spacers are schematically depicted under the charts as combs: green color teeth indicate matching-, while red color teeth indicates the presence of a mismatching nucleotide within the spacer; numbering of tooth position corresponds to the distance of the nucleotide from the PAM; the 20th nucleotide of the spacer is indicated as capital letter. Tukey-type notched boxplots by BoxPlotR[152]: center lines show the medians; box limits indicate the 25th and 75th percentiles; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles; notches indicate the 95% confidence intervals for medians; crosses represent sample means; data points are plotted as open circles and correspond to the different targets tested (in total 26 and 10, respectively, for A and B).

These results indicate that these increased fidelity nucleases are generally not compatible with the approaches examined here and can be routinely used only with perfectly matching 20G-sgRNAs. It is presumable from these results, and the lack of general alternative methods, that a new method that would relax the restriction for a 5'-end G nucleotide and would be compatible with the increased fidelity nucleases might prove to be a very valuable tool.

4.3. Cleavage activities of SpCas9 variants with 21G-sgRNAs and partially mismatching 20G-sgRNAs show positive Pearson correlations.

From the observed target cleavability ranking shown in the first chapter (Fig. 4), it should follow that the higher the cleavage activity of a nuclease with a mismatching spacer on a target, the higher is its activity with any other type of modified/imperfectly matching (5’ extended, truncated, or with another mismatching position) guide RNA on the same target, if they act using the same mechanism.

Such a correlation is discernible from our results. Correlation matrix analysis of the cleavage activity data shows positive Pearson correlations (0.83 or 0.78) that are significant between the disruption activities of either eSpCas9 or SpCas9-HF1 programmed with partially mismatching 20G-sgRNAs or 21G-sgRNAs targeting the same sites (Fig. 6A, C). In addition, differing positions of mismatches for the same targets also showed significant positive correlation (between 0.73 to 0.93 and 0.73 to 0.97 for eSpCas9 and SpCas9-HF1, respectively) (Fig. 6B, D). This result supports the idea that a

similar mechanism determines how these partially mismatching and 5’ modified sgRNAs affect the cleavage activities of SpCas9 nucleases.

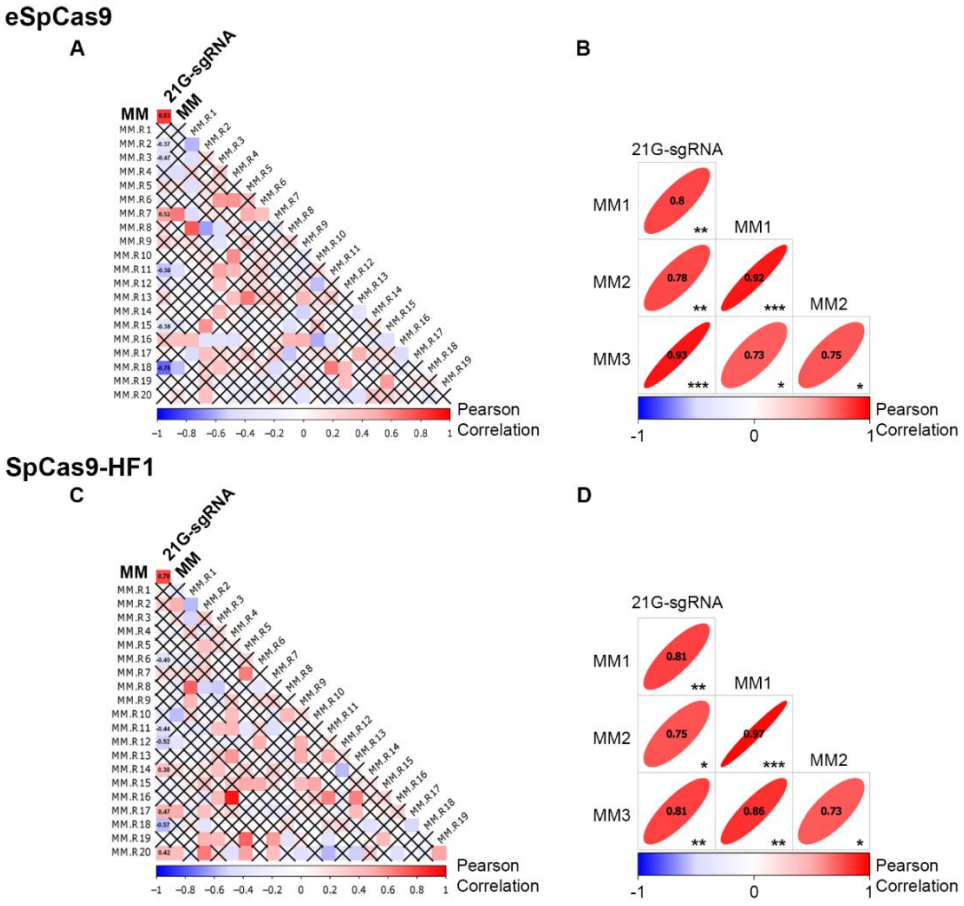


Figure 6. Correlation between the effects of mismatching 21G-sgRNAs and partially mismatching 20G-sgRNAs on the activities of eSpCas9 and SpCas9-HF1. The Pearson correlation was calculated by Elfrieda Fodor. EGFP disruption activity values normalized to WT of (**A, C**) eSpCas9 and (**B, D**) SpCas9-HF1 were analyzed when the nucleases are loaded either with 5' mismatching 21G-sgRNAs or with partially mismatching 20G-sgRNAs (“MM”, each target with 9 one-base mismatching sgRNAs for three different positions, each of the three possible mismatches as shown in Fig. 3A), which were examined either the 3 positions together (in average, A, C: MM) or each position separately (B, D: MM1, MM2 or MM3). Values are obtained on 10 target sites each with 9 partially mismatching 20G- and one mismatching 21G-sgRNAs. Numbers represent Pearson correlation coefficients; ellipses on C, D, are the 95% confidence intervals of the correlation coefficients. 2-tailed test of significance is used. ***: $p \leq 0.001$, **: $p \leq 0.01$, *: $p \leq 0.05$. **A, C,** Correlation tested between values corresponding to the same target ("21G-sgRNA" vs. "MM ") and when the mismatch-activity values are randomly shuffled (MM.R1-MM.R20) between targets. Insignificant values with $p>0.05$ are marked by X

at the corresponding positions; respective correlation values are added to the first columns. **B, D,** Correlation analysis on the activity values obtained within the three mismatching positions ("MM1-3") and the mismatching 21G-sgRNA of the same target, for the 10 targets.

4.4. Increased fidelity variants show closely WT-level binding activity on target sites cleaved inefficiently.

Another application of SpCas9, besides being a programmable nuclease, is using its inactive WT variant (dWT SpCas9), which retains its sequence-specific DNA binding activity, for delivering effector domains precisely to a chosen locus within the genome[11,126-133]. eSpCas9 and SpCas9-HF1 were designed that way to have a weaker interaction either with the non-targeted DNA strand or with the targeted DNA strand, respectively. Therefore, it was expected that they have reduced DNA binding compared to the WT SpCas9. We wondered how e- and SpCas9-HF1 and the higher fidelity HeFSpCas9 will perform in a transcriptional activation experiment which assumes DNA binding. For this reason, we compared their efficiency using the mix of five 21G-sgRNAs that generally diminishes their cleavage activity. The sgRNAs are targeting the promoter region of the *Prnp* gene that drives the expression of an EGFP cassette in a modified N2a cell line (N2a.EGFP). We found that all dead SpCas9 variants (deSpCas9, dSpCas9-HF1 and dHeFSpCas9) demonstrated activities comparable to that of the dWT SpCas9 level, resulting a 15-20-fold activation (Fig. 7A).

Another set up for transcriptional activation is when active SpCas9 is used, which binds the target DNA but without cleaving it when truncated sgRNAs are used. Although active WT SpCas9 is able to cleave on-target sites even with sgRNAs containing 17-19 nt-long truncated spacers[75,78,97], we can use it for sequence-specific binding without target cleavage if sgRNAs which contain only 14-15 nt-long spacers are used[103,153,154]. Our previous results showed that increased fidelity nucleases exhibit no cleavage activity when employing truncated sgRNAs missing more than two nucleotides (data not shown here, Kulcsár et al. Additional file 1: Figure S2f)[138]. Therefore we were curious how truncated sgRNAs would perform with active increased fidelity variants in a transcriptional activation experiment. We compared their efficiency using the mix of five truncated 15G- sgRNAs (in the same experimental set and same target sites as the ones in Fig. 7A). Contrary to the expectations, all increased fidelity nucleases demonstrated

comparable activities to WT, resulting a 15-20-fold activation (Fig. 7B), which is a similarly high level of activation compared to that what we see with 21G-sgRNAs in Figure 7A. These results, although indirect, suggest that the binding of the increased fidelity nucleases to their targets are not impaired even when the on-target activity is diminished with altered sgRNAs.

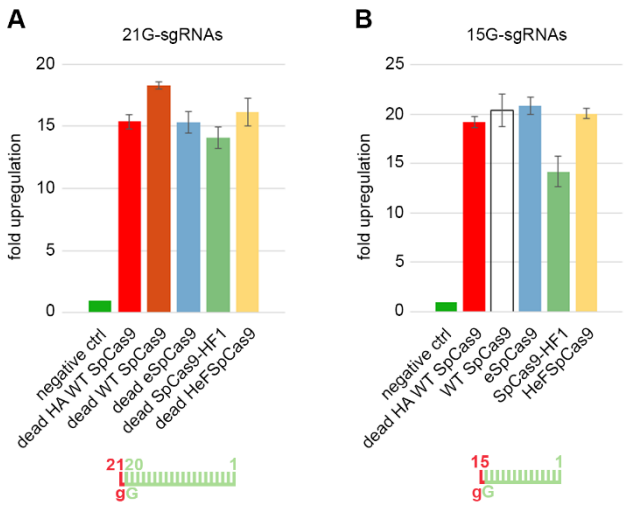


Figure 7. Transcription activation is not impaired compared to WT when the nuclease variants are charged with 5' modified sgRNAs. Transcription activation of a single copy of EGFP coding sequence inserted after the *Prnp* promoter, employing five target sites (same sites in case of both 15G- and 21G-sgRNAs) in the promoter region. **A,** Dead nuclease variants programmed with the mixture of five 21G-sgRNAs (20 nt-long MS2 aptamer-containing sgRNAs with an extra 5' G). **B,** active nuclease variants programmed with the mixture of five 15G-sgRNAs (14 nt-long truncated MS2 aptamer-containing sgRNAs with an extra mismatching 5' G). All SpCas9 variants were flanked with FLAG-tag, except dead HA-tag flanked WT SpCas9. Spacers are schematically depicted under the charts as combs, highlighting the main features of the spacers. Means are shown, error bars represent the standard deviation (s.d.) for n=3 biologically independent samples.

Employing a more direct *in vitro* approach to examine their DNA binding we performed polyacrylamide-gel electrophoretic mobility shift assay (EMSA) on the target DNAs by the nucleases. A serial dilution of the pre-incubated dWT SpCas9-sgRNA complex between the range of 32 to 4-fold of the molar amount of PCR amplified DNA target was tested to find the proper ratio where the system is not in saturation (data not shown here, Kulcsár et al.: Additional file 1: Figure S10a)[138]. Using the selected molar ratio first (SpCas9-sgRNA complex:target DNA = 4:1), we examined three targets that were shown to be cleaved efficiently by WT but not by HeFSpCas9. By these experiments, we confirmed that although HeFSpCas9 does not show nuclease activity, it retains most of its DNA-binding abilities to these targets (Fig. 8A). Furthermore, to better

understand the effect of appending an extra 5’ G nucleotide to the end of the sgRNAs we examined the in vitro binding of eSpCas9 and SpCas9-HF1 charged with matching 21G-sgRNAs to selected targets that were only cleaved when the 20G-sgRNAs were applied in the disruption assay (Fig. 5). We found that although the 5’ G extension fully diminishes the cleavage activities of these variants on these target sites in the GFP disruption assay, their binding seems to remain unaffected (Fig. 8B).

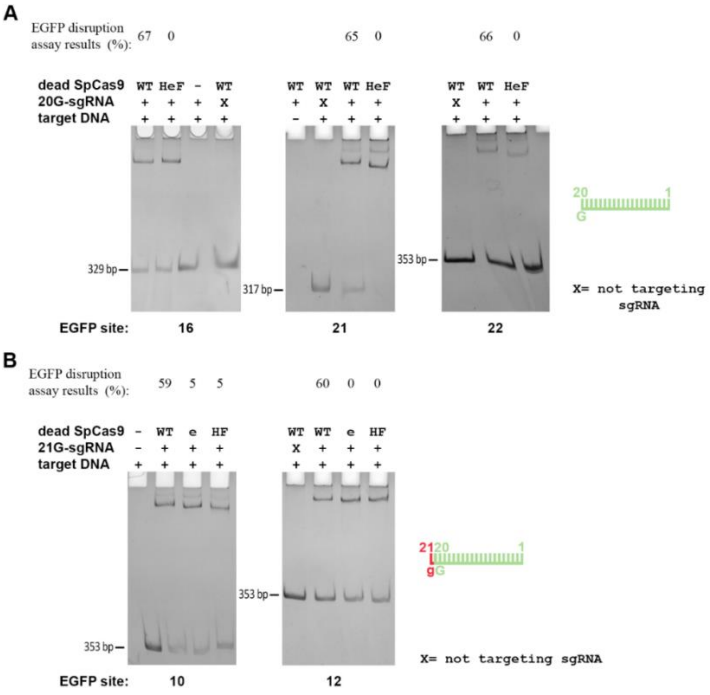


Figure 8. *In vitro* DNA binding of the increased fidelity SpCas9 nuclease variants to inefficiently cleaved targets. The EMSA experiment were carried out by Zoltán Ligeti. The SpCas9-sgRNA complex : target DNA molar ratio is 4:1 that is a sensitive condition to report on the binding activities**. A,** *In vitro* binding of dead WT and dead HeFSpCas9 charged with 20G-sgRNAs that are cleaved efficiently by WT SpCas9 but not by HeFSpCas9. X indicates sgRNAs that do not target the DNA. **B,** *In vitro* binding of dead eSpCas9 and dead SpCas9-HF1 charged with matching 21G-sgRNAs to selected targets that were only cleaved by the nuclease variants when the same but only perfectly matching 20G-sgRNAs were applied in the disruption assay. Spacers are schematically depicted beside the charts as combs, highlighting the main features of the spacers.

4.5. Blackjack-SpCas9-HF1 works with 21G-sgRNAs.

According to the cleavage competent structure, one explanation for the diminished cleavage activity when employing 21G-sgRNAs could be that it is the result of a capping of the 5’ end of the sgRNA by two amino acids, namely Glu1007 and Tyr1013, which are connected via a surface loop (Fig. 9)[66,155]. We proposed that removing this cap by mutations would allow for a 5’ extension of the sgRNA and it could be achieved without

disruption of the structural features of the folded polypeptide chain. Such modification would allow the increased fidelity nucleases to work with similar efficiency when charged with either 20G-sgRNAs or 21G-sgRNAs, thereby increasing their available target space. It has been shown that a 5' GG dinucleotide extension of the sgRNA increases the fidelity[74]. We hypothesized that this fidelity increment was caused mainly through the disruption of the stabilizing cap-interaction by the 5' GG extension and therefore removing the cap by mutations may also increase the fidelity of the nuclease.

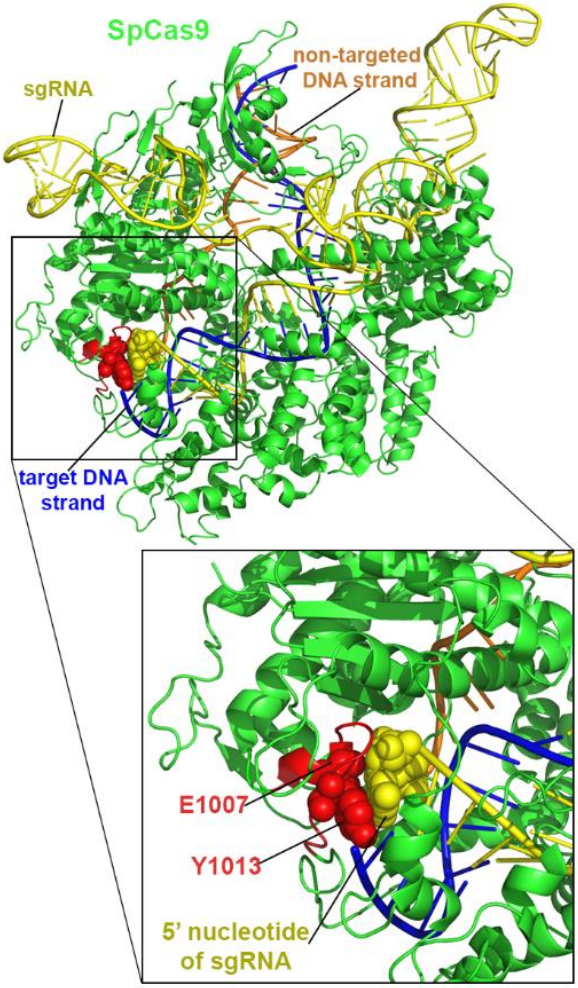


Figure 9. Structure-guided mutagenesis increases on-target activity of SpCas9-HF1 with 21G-sgRNAs. X-ray crystallography derived structure of SpCas9-sgRNA-DNA complex in the conformation closest to the cleavage competent state (PDB ID: 5f9r)[156].

We chose SpCas9-HF1 from among the high-fidelity nucleases as a starting platform and generated a mutant by replacing both Glu1007 and Tyr1013 with glycine to eliminate the presence of sidechains at these positions. In addition, we generated two deletion mutants within the region from 1004 aa. through 1014 aa. (positions where the remaining ends of the polypeptide chain seemed to be connectable without causing major distortions of the protein structure) either by completely removing this segment or by replacing it with two adjacent glycine residues, in order to eliminate the loop (Fig. 9). Interestingly, both variants containing the deletions were active with 21G-sgRNAs when tested in an EGFP-disruption assay, but not the glycine substitution-mutant (data not shown). In consequence, we decided to proceed further with deletion variants. We created 16 further deletion mutants in that region, specifically between aa. 1003 and aa. 1017, by

completely removing or exchanging segments of various lengths harboring the loop, with between one and four amino acids with no or small side chains, and tested them on 13 target sites both with 20G- and 21G-sgRNAs (data not shown). None of them exceeded the previously characterized variant: containing only two glycine residues between the amino acids L1004 and K1014, what we named Blackjack-SpCas9-HF1 (B-SpCas9-HF1; Blackjack mutations designated by the "B-" prefix) (Fig. 10A). This candidate was considered to be the one exhibiting the highest on-target activity with 20G-sgRNAs and demonstrating the highest improvement with 21G-sgRNAs.

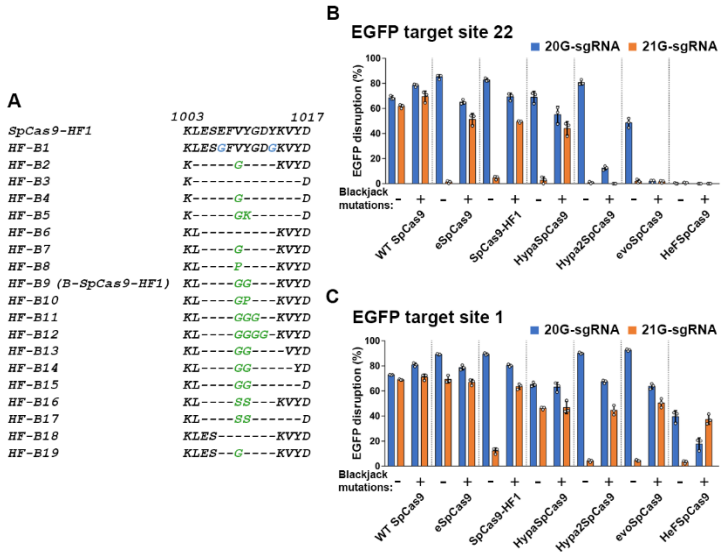


Figure 10. HF-B9 "Blackjack" candidate proved to be the optimal variant. A, Amino acid sequences between residues 1003 and 1017 of the SpCas9-HF1 and Blackjack SpCas9-HF1 candidates examined. Deletions (-), insertions (green) and substitutions (blue) are indicated. **B-C,** Blackjack mutations increase on-target activities of increased fidelity variants a target dependent manner with 21G-sgRNAs. Means are shown, error bars represent the standard deviation (s.d.) for n=3 biologically independent samples (overlaid as white circles).

4.6. The Blackjack increased fidelity variants

Next, we introduced the Blackjack mutations into all six increased fidelity variants and into the wild type nuclease and compared their on-target activities with 20G- and 21G-sgRNAs to see whether they retain their 20G-sgRNA level activity with 21G-sgRNAs. Blackjack mutations increase the on-target activity of the variants with 21G-sgRNAs up to 17-folds. The results obtained with two representing targets with different contributions are presented in Figure 10B and C. Due to the different target contribution (EGFP target site 1 from the *high-contributing* range and EGFP target site 22 from a

lower range) the various SpCas9 variants cleave the target sites with different efficiency. This can be seen not only with 20G- but with 21G-sgRNA cleavage as well.

When the activity of the 14 SpCas9 variants (both Blackjack and their "parental" variants) are checked with 21G-sgRNAs on those EGFP targets out of the 47 where the Blackjack variants retain their on-target activities with 20G-sgRNAs, all Blackjack variants exhibit greatly increased activities compared to their parental variants (Fig. 11A).

We were curious if Blackjack mutations indeed increase the fidelity/target-selectivity ranking of these variants. We compared their on-target activities on the same 47 targets. The pattern in Figure 11B clearly shows that on average the Blackjack mutations increase the target-selectivity (as discernable from the decreased average on-target activity) of all SpCas9 variants except that of the wild type. Based on our earlier results this suggest that their fidelity is increased as well. We carried out a mismatch screen on the same 16 target sequences using the 144 mismatching sgRNAs as earlier in Figure 4B. The results demonstrated that introduction of the Blackjack mutations raises the rank of SpCas9 variants by increasing their fidelity (Fig. 11C; B-WT and WT SpCas9 data not shown).

To validate these conclusions and see if the disruption of the cap interaction indeed increases the genome-wide fidelity of WT SpCas9 we applied GUIDE-seq - an unbiased genome-wide off-target analysis method[75] - to six target sites. Both the 5' G extension and the Blackjack mutations decreased the number of off-target events detected and increased the percentages of on-target vs. off-target reads suggesting the fidelity increasing effect in line with our earlier results (Fig. 11D and App. Fig. 2).

We interpreted these results to mean that the Blackjack mutations successfully removed the cap from the 5' end of the sgRNA in the cleavage-competent conformation of the SpCas9-sgRNA-DNA complex allowing the effective use of 21G-sgRNAs with increased fidelity variants. However, the mutations also seem to alter the protein conformation in such a way as to increase the fidelity/target-selectivity of the variants. The new Blackjack variants lie within the ranges of their parental variants (i.e. … < eSpCas9 < SpCas9-HF1 < **B-eSpCas9** < HypaSpCas9 < **B-SpCas9-HF1** < …). As a result of this "up-shifted" fidelity from the Blackjack mutations there is no Blackjack-SpCas9 variant that is optimal to targets in the lower cleavability ranks (such as eSpCas9, SpCas9-HF1) in the present repertoire.

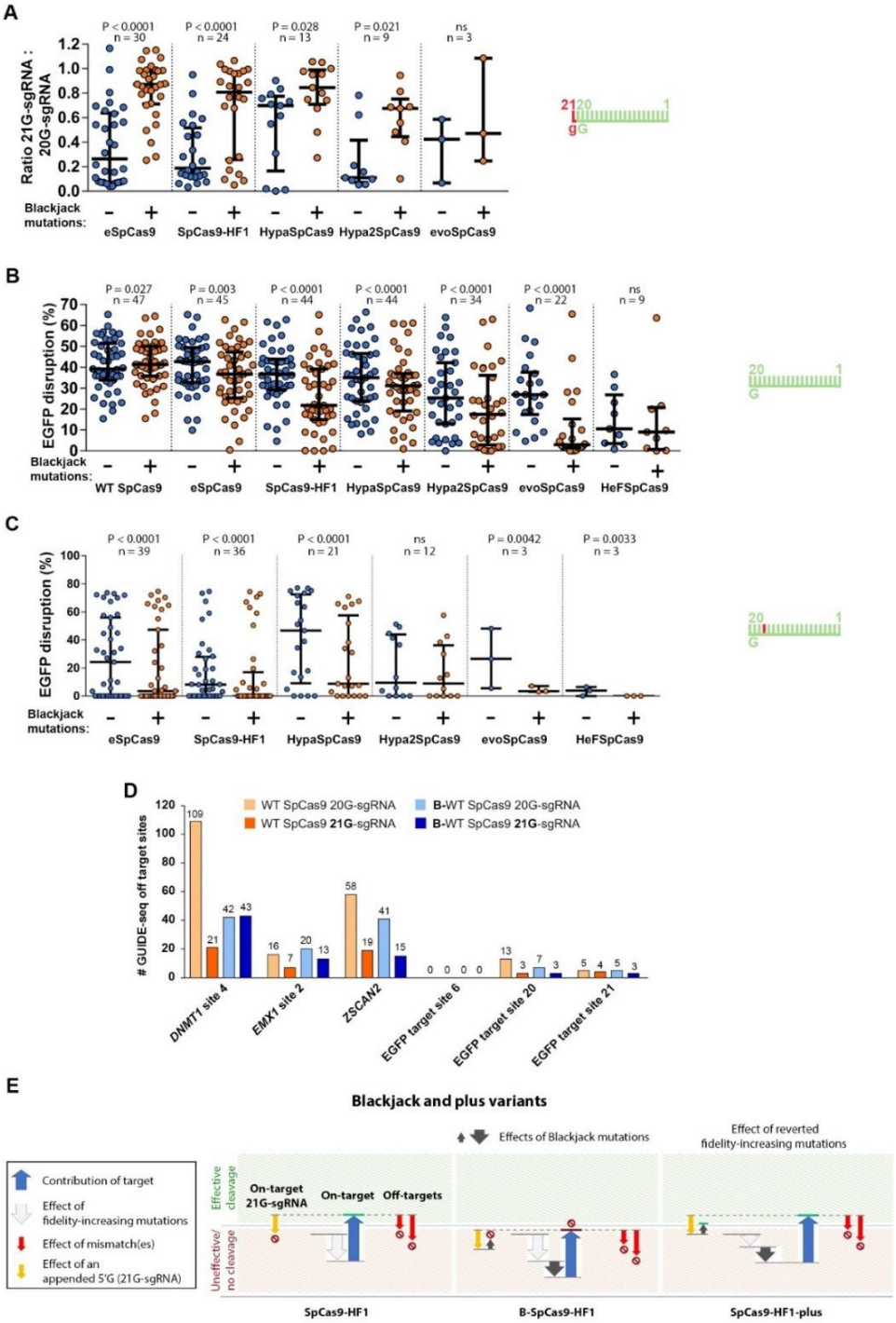


Figure 11. The Blackjack mutations increase not only the activity of increased fidelity nucleases charged with 21G-sgRNAs but their target-selectivity and fidelity. A, On-target activities with 21G-sgRNAs for which the SpCas9 variant without Blackjack mutations exhibits at least 70% on-target activity compared to WT SpCas9. **B,** EGFP disruption activities with perfectly matching 20G-sgRNAs. Results are shown only for those target sites where the SpCas9 variant without Blackjack mutations exhibits higher than background level cleavage. **C,** EGFP disruption activities with partially mismatching 20G-sgRNAs. Results are shown only for those

target sites where both the non-Blackjack parent- and Blackjack-SpCas9 variant exhibits at least 70% on-target activity (with perfectly matching 20G-sgRNA) compared to WT SpCas9. The sample points in (**B**) and (**C**) correspond to targets and sgRNAs presented in Figure 4A and B, respectively. (**A-C**) Spacers are schematically depicted beside the charts as combs, highlighting the main features of the spacers. The median and the interquartile range are shown; data points are plotted as open circles representing the mean of biologically independent triplicates. Statistical significance was assessed using Paired-samples Student's t-test or Wilcoxon signed ranks test as appropriate; ns: not significant. **D,** Bar chart of the total number of off-target sites detected by GUIDE-seq for WT and B-WT SpCas9 variants on six target sites targeted with 20G- or 21G-sgRNAs. **E,** Schematics to explain the rational for the generation of the *plus* variants: An appended G nucleotide (yellow arrow) similarly to fidelity-increasing mutations diminish cleavage activity of SpCas9-HF1. Blackjack mutations exert two opposite effects: increase fidelity/target-selectivity (black down arrow) and diminishes the effects of a 5' appended G (black up arrow). By reverting a fidelity-increasing mutation back to the WT the composite effect of the fidelity-increasing mutations (grey arrow) and Blackjack mutations (black down arrow) on the activity of SpCas9-HF1-*plus* became the same as with SpCas9-HF1 (grey arrow) but preserve the diminished effect of a 5' appended G seen with B-SpCas9-HF1.

4.7. e-SpCas9-*plus* and SpCas9-HF1-*plus* variants

To have a more complete target coverage, we decided to create Blackjack variants that have identical fidelities/target-selectivities to those of eSpCas9 and SpCas9-HF1. We proposed that we could achieve this by first restoring some of the 'parental' fidelity-increasing mutations to their wild type residue variants - due to the cumulative effect of different mutations examined in this study - that presumably would decrease the fidelity of Blackjack variants without effecting their 21G tolerance (Fig. 11E). eSpCas9 possesses three mutations (K848A, K1003A, R1060A) while SpCas9-HF1 possesses four (N497A, R661A, Q695A, Q926A). After examining the data in the studies describing their development[103,104], we constructed four and seven candidates from B-eSpCas9 and B-SpCas9-HF1, respectively, lacking one or two 'parental' mutations at a time (Fig. 12A, D). We selected those residues for which we conjectured that their contributions to the increased fidelity of the respective nuclease would be comparable to those of the Blackjack mutations.

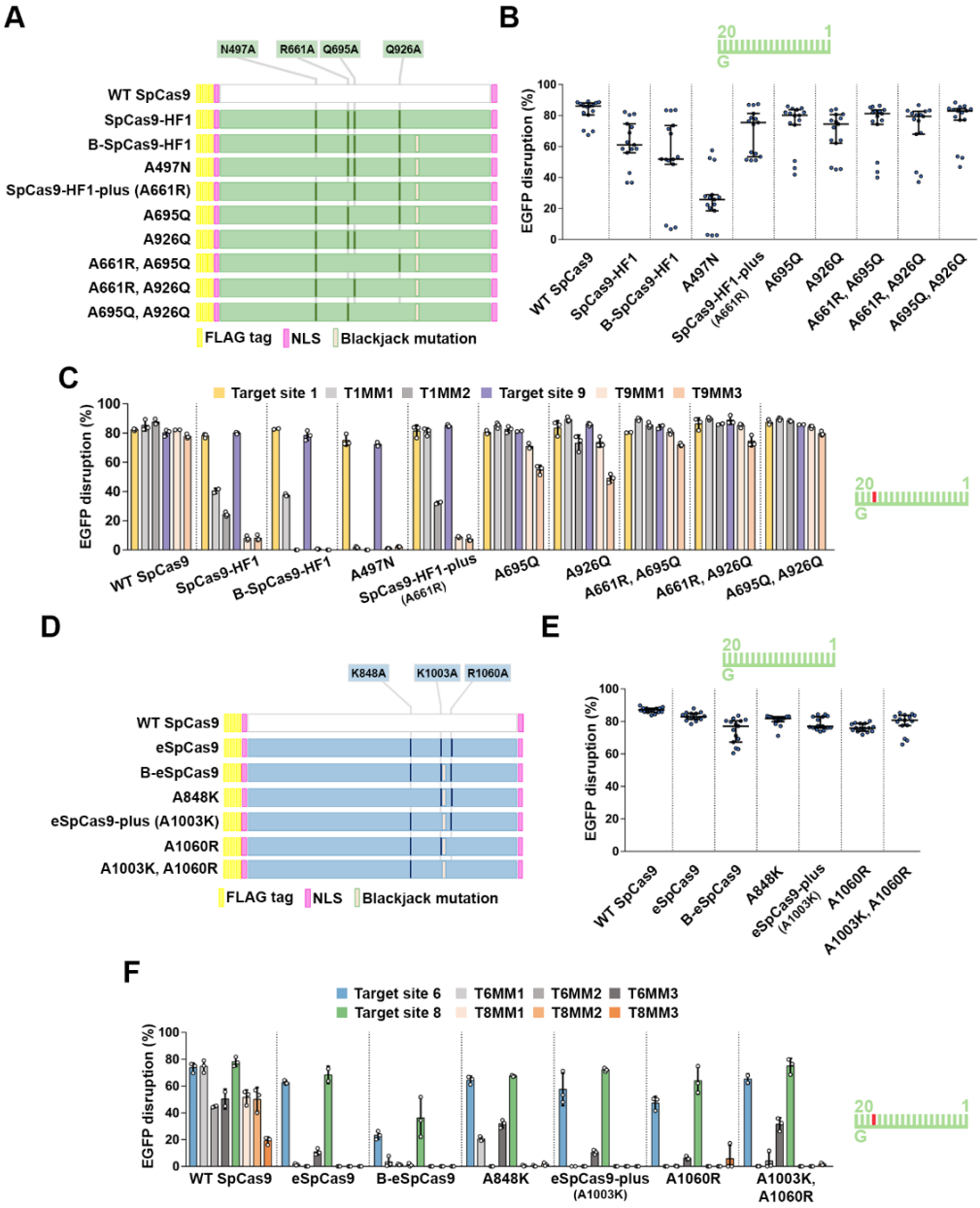


Figure 12. Restoring mutations to wild type amino acids lowers the target-selectivity and the fidelity of B-eSpCas9 and B-SpCas9-HF1. **A, D,** Schematic representation of the mutations in each variant examined. **B, E,** On-target activities using 20G-sgRNAs measured on 5 target sites (n=3), employing EGFP disruption assay, median and interquartile range are shown. **C, F**, Mismatch screen results from EGFP disruption assay. Target sites and matching (e.g., T1, T6) or mismatching sgRNAs (e.g., T1MM1, T6MM1) are the same as in Figure 4A and B. Schematics highlighting the main features of the spacers used are added to the charts. Means are shown, error bars represent the standard deviation (s.d.) for n=3 biologically independent samples (overlaid as white circles). Spacers are schematically depicted beside the charts as combs, highlighting the main features of the spacers.

For SpCas9-HF1 we picked 5 targets on which it has considerable activity employing 20G-sgRNAs and, on which B-SpCas9-HF1 exhibits strongly decreased activities with 21G-sgRNAs due to its increased target-selectivity and tested the residue-reverted variants. Except the candidate in which A497 was reverted (surprisingly, it seems to show decreased on-target activities), all new candidates exhibit increased on-target activity with 20G-sgRNAs compared to B-SpCas9-HF1 (Fig. 12B). This suggests that the target-selectivity obtained with revertants was successfully reduced and their fidelity correspondingly lowered compared to B-SpCas9-HF1 except for the A497N reversion. To find the variant whose fidelity most closely matches that of SpCas9-HF1 we employed mismatching sgRNAs to two selected targets (Fig. 12C). The reversion of A661 resulted in a similar fidelity as SpCas9-HF1. We named this Blackjack variant as *SpCas9-HF1-plus*. For B-eSpCas9, we proceeded similarly as for B-SpCas9-HF1 to create the revertants and we found that the reversion of A1003 resulted in the closest fidelity match to that of eSpCas9's (Fig. 12E, F). We named this Blackjack variant *eSpCas9-plus*.

Western blotting indicated that Blackjack mutations do not alter the expression level of SpCas9 and the amounts of the *plus* variants expressed at steady state are comparable to their parent variants (Fig. 13A). We compared the selected *plus* variants' on-target activities with 20G-sgRNAs on 25 EGFP targets. Both eSpCas9-*plus* and SpCas9-HF1-*plus* reached the on-target activities of their original counterpart variant (eSpCas9 and SpCas9-HF1, respectively) on this set of target sequences (Fig. 13B). To challenge the *plus* variants when checking their activities with 21G-sgRNAs, different sets of 10 targets were assayed with e- and SpCas9-HF1 to exploit targets on which the parent nucleases exhibited strongly decreased on-target activity with 21G-sgRNAs (but WT-level activity when 20G-sgRNAs are utilized). Since *plus* variants' target-selectivity/fidelity is higher than that of WT SpCas9, they are not expected to reach WT-level activity with 21G-sgRNAs (neither with 20G-sgRNAs). Figure 4A shows that the original counterparts, eSpCas9 and SpCas9-HF1 exhibit slightly decreased on-target activities with 20G-sgRNAs, 93% and 82% relative to WT SpCas9 in average, respectively. This relative activity level should be reached by the *plus* variants if they worked with 21G-sgRNAs with identical efficiency as the corresponding original variants do with 20G-sgRNAs. Indeed, with 21G-sgRNAs they did demonstrate 89% and 82%; in contrast, their parent variants demonstrate only 10% and 16%, respectively (Fig. 13C, D).

We further examined the on-target activities of the *plus* variants in HEK293 cells, monitoring their indel-inducing activities by NGS. We selected 23 target sites from the human *FANCF* and *VEGFA* loci. Sixteen of them can be targeted with 21G-sgRNAs and seven with 20G-sgRNAs. The on-target activities of the *plus* variants with the seven 20G-sgRNAs match those of their corresponding parental variants. However, with 21G-sgRNAs they show much higher on-target activity, similar to that what we would expect from e- and SpCas9-HF1 with 20G-sgRNAs (data not shown).

To compare the fidelity of the *plus* variants with their parental ones, 13 targets were selected (39 positions were examined, each target at three positions by applying the mix of three-three mismatching sgRNAs, such as in earlier screens). The off-target activity of eSpCas9-*plus* resulted in an identical off-target-cleavage pattern, matching the fidelity of eSpCas9 (Fig. 13E). The off-target activity of SpCas9-HF1 and SpCas9-HF1-*plus* compared in a similar way also gave rise to very similar patterns, closely matching each other's fidelities (Fig. 13F). These data demonstrate that the *plus* variants possess identical fidelity but strongly increased on-target activity with 21G-sgRNAs, and thus, can be utilized on a much wider range of targets for high fidelity editing compared to their enhanced and high-fidelity parent variants.

We also wanted to compare the fidelity of these nucleases by GUIDE-seq[75]. We selected 7 target sites that can be targeted by 20G-sgRNAs to make sure that not just the nuclease variants containing Blackjack mutations are able to cleave the on-target sequences. Among these targets, three (*VEGFA* site 2, HEK site 4, *FANCF* site 2) have been used to characterize the off-target activities of the increased fidelity variants in earlier studies[104,106,107,121]. As expected, we found that all four increased fidelity variants demonstrate greatly increased fidelity compared to the wild type protein and behave similarly to each other on six out of the seven targets. On *VEGFA* site 2 eSpCas9-*plus* cleaves less, while SpCas9-HF1-*plus* more off-target sites compared to their parental variants (Fig. 13G and App. Fig. 3). These results are consistent with the contention that the fidelity of these *plus* variants matches that of their original parental (non-Blackjack) counterparts.

We were also curious to compare the on-target activities of the *plus* variant with different sgRNA processing approaches (which can be used to target non-20G-N19-NGG target sites as well) to see whether the *plus* variant offers advantages in those applications

where these other approaches can also be applied. Testing on 19 target sites, the *plus* variant showed higher activities with 21G-sgRNAs than either the tRNA-[150] or ribozyme-flanked sgRNAs[146,157] with eSpCas9 (Fig. 13H, for target and spacer sequence information see App. Table 9). Some target sequences in case of tRNA- and ribozyme-flanked sgRNAs were cleaved with reduced efficiency by WT SpCas9, suggesting that the understanding of the sequence dependencies of these approaches needs a more comprehensive investigation.

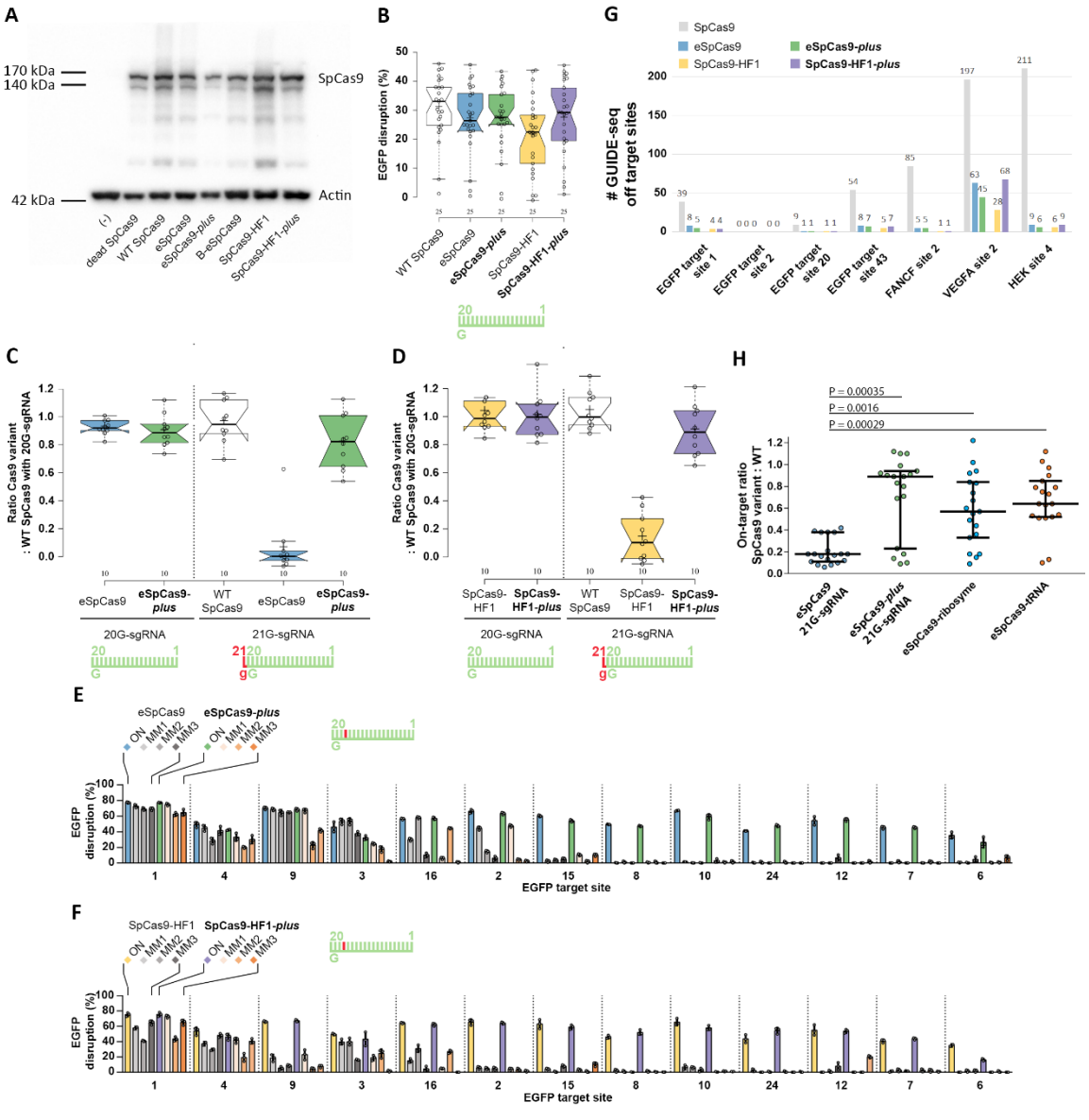


Figure 13. Characterizing eSpCas9-*plus* and SpCas9-HF1-*plus* variants. A, Immunoblot analysis of the expression levels of SpCas9 nuclease variants (~160 kDa) in cell lysates of reporter N2a.dd-EGFP cells transfected with the indicated nuclease constructs. β-actin (~42 kDa) was used

as a control for total protein amounts analyzed. **B-F,** EGFP disruption activity **B**, with 20G-sgRNAs targeting 25 sites; **C, D,** with either 20G- or 21G-sgRNA pairs targeting two alternative sets of 10 different sequences shown as the ratio of variant activity to WT activity; **E, F,** with 20G-, perfectly matching or partially mismatching sgRNAs targeting 13 sites; sites are shown in the order of decreasing cleavability according to the ranking shown in Figure 4A**.** Spacers are schematically depicted beside the charts as combs, highlighting the main features of the spacers. **B-D,** Tukey-type boxplots by BoxPlotR[152]: center lines show the medians; box limits indicate the 25th and 75th percentiles; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles; notches indicate the 95% confidence intervals for the medians; crosses represent sample means; data points are plotted as open circles representing the mean of biologically independent triplicates. **E, F,** Means are shown, error bars represent the standard deviation (s.d.) for n=3 biologically independent samples (overlaid as white circles). **G,** Bar chart of the total number of off-target sites detected by GUIDE-seq for SpCas9 variants on seven sites targeted with 20G-sgRNAs. **H,** EGFP disruption activities of SpCas9 nuclease variants programmed with 5' extended 21G-sgRNAs, hammerhead ribozyme flanked-sgRNAs[146] or rice tRNA flanked-sgRNAs[150] targeting 19 EGFP target sites (20th position non-G) shown normalized to WT SpCas9. The median and the interquartile range are shown; data points are plotted as open circles representing the mean of biologically independent triplicates. Statistical significance was assessed by the Wilcoxon signed rank test.

4.8. RNP form of Blackjack and *plus* variants increase specificity.

The development of two new increased fidelity variants, Sniper- and HiFi SpCas9 has been reported more recently, claiming they work effectively in RNP form, and Sniper SpCas9 is able to work even with 5' modified sgRNAs, unlike former increased fidelity variants[120,121]. We compared the on-target activity and fidelity of these two variants to eSpCas9 to understand their extra properties. Both variants showed similar target-selectivity as WT SpCas9 and higher fidelity compared to that (App. Fig. 4). However, they had lower fidelity compared to the former "low" increased fidelity variants such as eSpCas9. The fact, that Sniper SpCas9 being closest in ranking to WT offers an explanation for its ability to work with 5' modified sgRNAs. However, it is not clear why they possess high activity in RNP form, while the former increased fidelity variants have been reported that they show a strongly reduced activity in RNP form[120].

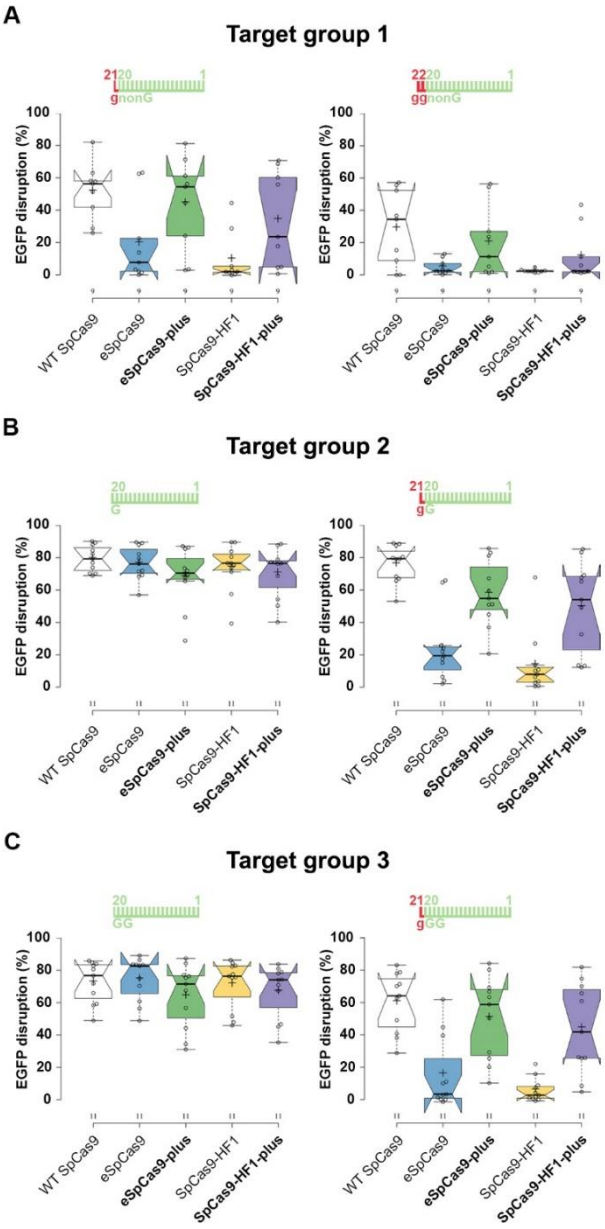


Figure 14. The *plus* variants are effective when transfected as preassembled RNP form. *In vitro* transcription and protein purification were carried out by Zoltán Ligeti and transfection by András Tálas. **A-C,** EGFP disruption assays. Target sequences start with 5' non-G-, G- or GG-nucleotides. Tukey-type boxplots by BoxPlotR[152]: center lines show the medians; box limits indicate the 25th and 75th percentiles; notches indicate the 95% confidence intervals for the medians; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles; crosses represent sample means; data points are plotted as open circles representing the mean of biologically independent triplicates. Spacers are schematically depicted beside the charts as combs, highlighting the main features of the spacers.

RNPs are the method of choice for prospective clinical applications, and we wondered if Blackjack variants are able to provide optimal high fidelity editing for the majority of the targets that require nucleases with higher fidelity compared to Sniper or HiFi SpCas9. Thus, we selected 31 sequences to assay for EGFP disruption by eSpCas9 and SpCas9-HF1 and by their *plus* variants delivered in RNP form. Since it seems that there is no consensus about the requirement of the T7 polymerase for the preferred starting sequences of the transcript, we selected sequences that start with non-G, G or GG nucleotides and we targeted them systematically with *in vitro* transcribed, 5' G- or GG-extended or fully matching 20G-sgRNAs, as depicted in Figure 14A-C. Surprisingly, in contrast to that reported earlier[120], all variants show similarly high activities with 20G-sgRNAs as the WT protein (Fig. 14B - left boxplot). By contrast, only the *plus* variants demonstrate high activities with 21G-sgRNAs in pre-assembled RNP form as well,

reaching up to 23-fold higher activities than their parental variants (Fig. 14A-C). Thus, we concluded that *plus* variants are effective in the RNP form and together with the other Blackjack variants they provide high-fidelity editing in the whole target ranges of WT SpCas9 with both 20G- and 21G-sgRNAs and allow the effective use of *in vitro* transcribed 21G-sgRNAs.

Since increased fidelity variants are not compatible with the use of 5' truncated or modified sgRNAs – which have been shown to lower the off-target activity[74,75,78,97]-, their specificity cannot be further increased by these approaches. It has been reported that RNP delivery increases the specificity of genome modifications compared to that carried out on the same targets by WT SpCas9 but delivered as plasmids[96]. We were curious to see whether we could confirm the implication of the above results, namely that increased fidelity variants (including Blackjack and *plus* variants) in RNP form further increase the specificity of genome modifications compared to that carried out on the same targets by the same optimized increased fidelity nuclease variants but expressed from plasmids. We applied the GUIDE-seq method to detect genome-wide off-target events targeting three sequences, the EGFP site 20 for which both eSpCas9 and SpCas9-HF1 are expected to be optimal, and the EGFP site 43 and *FANCF* site 2 targets for both of which HeFSpCas9 is expected to be optimal (Fig. 15, App. Fig. 3). Among the Blackjack variants B-evoSpCas9 matches most closely the fidelity/target-selectivity of HeFSpCas9 based on the on-target and mismatch screen, therefore, we chose to use this for EGFP site 43 and *FANCF* site 2 targets. To demonstrate the advantage Blackjack variants can provide, we applied them using 21G-sgRNAs when RNP form was used. The highest specificity gene editing is detected with all three target sequences when employing them in RNP form in contrast to employing them in plasmid form. The specificity improvement with the *FANCF* site 2 sequence is particularly remarkable, which has been used in other studies to challenge almost all increased fidelity nucleases[104,106,107]; however, none of them could cleave it without considerable off-target cleavage as also confirmed in this study. Among all of the non-Blackjack increased fidelity nucleases HeFSpCas9 shows optimal specificity for the *FANCF* site 2 with least off-target cleavage (2417 on-target vs. 15 off-target reads). The closest Blackjack variant, B-evoSpCas9 in RNP form cleaved the target (7896 on-target reads) without any detectable off-target cleavage (Fig. 15): demonstrating the power and potential of using increased fidelity variants in combination with the RNP

format (in case of *plus* and Blackjack variants even with 21G-sgRNAs) for further increasing the specificity of genome modifications.

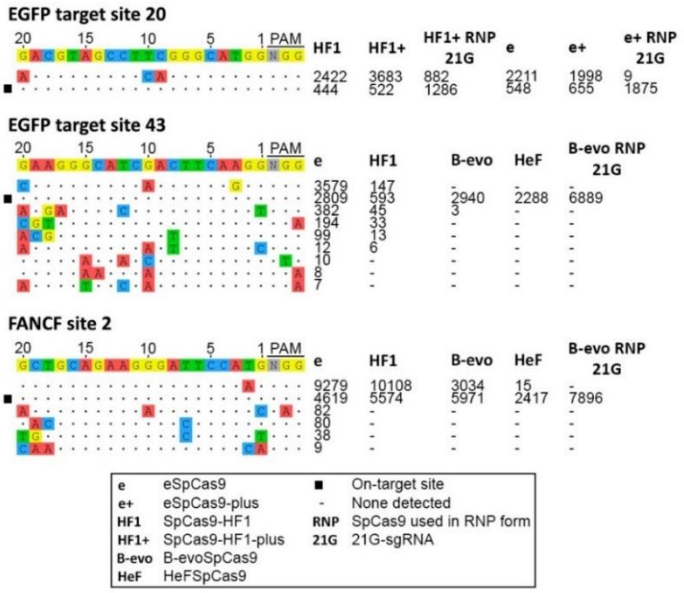


Figure 15. Blackjack variants of SpCas9 in RNP form further increase the attainable specificity of genome modifications. Off-target cleavage sites determined by GUIDE-seq. Read counts represent a measure of cleavage frequency at a given site; mismatched positions within the spacer or PAM are highlighted in color.

4.9. An example for the application of *plus* variants

Finally, we wanted to test the usefulness of *plus* variants in a practical application by presenting how inserting the EGFP sequences downstream of the mouse Shadoo promoter exploiting NHEJ. Five NGG PAM sequences are available at relevant positions in the gene but none of them are targetable with 20G-sgRNAs, presenting a good example where Blackjack variants can offer a specific advantage for an endeavor. We pre-screened the available five targets with several increased fidelity nucleases using a HR mediated integration assay. Since we have shown that using a 'self-cleaving' EGFP-expression plasmid or inserting a homologous recombination donor plasmid results comparable integration levels in case if the same sgRNA was used (data not shown, Tálas et al. Supplemnetary Fig S6)[137], we were able to choose the optimal targets and nuclease variants identified by the HDR pre-screening (Fig. 16A). We used these selected 'sgRNA – increased fidelity variant' combinations for the generation of the desired transgenic lines using the 'self-cleaving' donor plasmid (Fig. 16B). These results confirmed the advantage of the *plus* variants and how they could be used for targeting the whole target range what is available for WT SpCas9 but with lowered off-target effect.

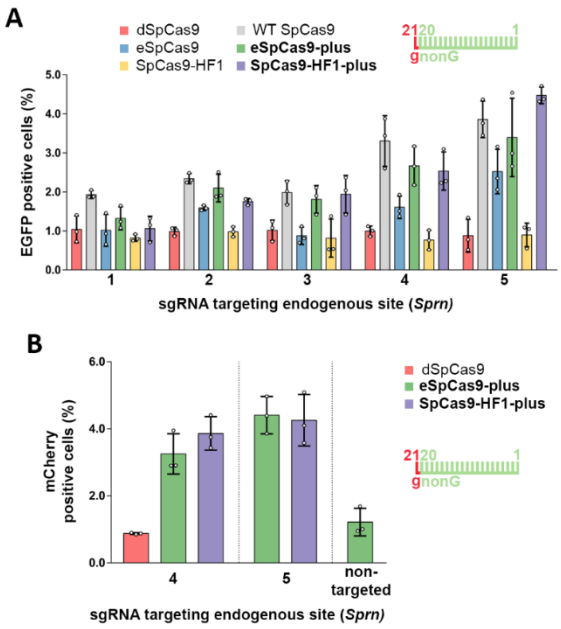


Figure 16. *plus* variants facilitate modification at the 5' coding region of the endogenous Shadoo (Sprn) gene. A, Increased fidelity nucleases complexed with 21G-sgRNAs were pre-screened for efficiency on 5 selected *Sprn* targets by the integration of a HR donor CMV-EGFP cassette. **B,** Based on results obtained from the prescreen on panel (**A**) *Sprn* target 4 and 5 were chosen for generating transgenic cell lines. *plus* variants were transfected along with corresponding 21G-sgRNAs and a 'self-cleaving' promoterless-EGFP-expression plasmid, which has to be integrated in-frame for *Sprn* promoter driven EGFP expression. **A, B,** Spacers are schematically depicted beside the charts as combs, highlighting the main features of the spacers. Means are shown, error bars represent the standard deviation (s.d.) for n=3 biologically independent samples (overlaid as white circles).

5. Discussion

Although there are different approaches that have been successfully used to decrease off-target cleavage activities of SpCas9, currently the most promising approach to decrease its off-target activity are the generation of increased fidelity variants. It has been shown that all these published variants show higher fidelity compared to WT SpCas9, but their systematic comparison was still missing.

One of the most important outcomes of our study is the characteristic pattern associated with the relative cleavage activities regarding both the on-target screen and mismatch screen of the 14 SpCas9 variants (Blackjack and their parental variants) that revealed their fidelity and target-selectivity rankings and the cleavability ranking of the targets as well.

The process of target recognition and cleavage by SpCas9 includes at least two major steps. First, the RNP complex binds to its target sequence in an inactive conformation state directed by the complementarity between the sgRNA and the PAM proximal region of the target site[158]. At the second step, the RNP complex proceeds to a cleavage competent, active conformation, depending on the complementarity of the sgRNA[106,158]. Mismatches between the sgRNA and the target in the PAM proximal region tend to decrease the dwelling time of the complex on the target sequence, abrogating binding, while mismatches in the PAM distal region decrease the time spent in the cleavage competent conformation[57,106,159]. Mutations in eSpCas9 and SpCas9-HF1 were shown to decrease the time the protein spends in the active, cleavage competent conformation, however, to different extents leading to a decreased *in vivo* cleavage rate with mismatching, but not with matching target sites and providing a mechanistic explanation for the higher fidelity of eSpCas9 and SpCas9-HF1[106].

In combination with the former mechanistic knowledge on the recognition and cleavage mechanism of SpCas9 variants[106,159,160], the rankings and other findings recognized by us shed light upon several important features of the process including the major factors that determines cleavage activities. Specifically, (1) all the variants tested here fit the fidelity and target selectivity ranking, suggests that the protein mutations (including Blackjack mutations) in case of any increased fidelity variants exert their effect similarly to that of eSpCas9 and SpCas9-HF1, regardless of the rationale for their

development, or whether being located in the REC3, HNH or RuvC domains, or whether they are in direct contact with the nucleic acids of the SpCas9-target DNA complex or not (Fig. 4). The generation of the *plus* variants further supports this rationale (Fig. 11). (2) Second, we demonstrated that PAM distal mismatches between the targets and the sgRNA affect target cleavage via the same mechanism as the protein mutations of increased fidelity variants do, a conclusion also supported by smFRET experiments[106]. The various effects of mismatches and protein mutations are at least partially additive in terms of shifting the conformational equilibrium toward the inactive conformation in case of a given target, which finally may diminish the cleavage rate (Fig. 4B, C). (3) Third, increased fidelity variants show decreased activity with 5' extended sgRNAs as well (Fig. 5 and 11A), suggesting a similar mechanism of action between 5’ extending (or truncating) the guide and the PAM distal mismatches in the sgRNA. This contention of similar mechanisms is further strengthened by the correlations we found between the activity-reducing effect of appending a 5' G to the sgRNAs and of using mismatching sgRNAs with increased fidelity nucleases (Fig. 6). Indeed, these have been shown for truncated sgRNAs[159] and confirmed in a recent pre-print study for 5’ G extended sgRNAs (https://www.biorxiv.org/content/10.1101/642223v1). Based on these observations we speculate that all kinds of alterations to the SpCas9/target DNA complex which destabilize the active and/or stabilize the inactive conformations could make the protein spend less time in the cleavage-competent, active conformation, and thus increase fidelity without a major effect on the target DNA binding. These alterations, beyond the protein mutations and target/sgRNA mismatches, may also include truncated sgRNAs[78] or sgRNAs containing modified bases[161] or DNA nucleotides[162]. Such alterations can either decrease the stability of the DNA-RNA hybrid helix embedded in the cleavage competent conformations or can stabilize interactions of the unwound strands of the hybrid helix in the inactive conformations as suggested recently by MD simulations[163]. (4) Fourth, the results seen in Figure 4 indicate that the efficiency and the specificity of target-cleavage by the increased fidelity variants are interrelated and highly determined by the nature of the target sites which can be ranked as well. These patterns observed also suggest a specific contribution of the target sequences, which may affect the inactive-active conformational transition step through the same mechanism (Fig. 4C). Altogether, these observations suggest that the contribution of the target sequences, the effect of the

mutations that cause increased fidelity (including mutations that reduce non-sequence specific protein-DNA contacts and Blackjack mutations) and the effect of the sgRNA modifications (mismatches between the sgRNA and the target DNA, 5' modifications) are apparently cumulative (Fig. 11E) and they are affecting the catalytic activity of the cleavage complex without much altering the target binding of SpCas9 (Fig. 7, 8). The above mentioned effects offer an explanation for both the results seen in our experiments and the characteristic pattern seen in Figure 4A and 4B and are in line with other studies[103,104,106,107,121,146,151].

These results therefore indicate that generating a superior SpCas9 which has high specificity, yet is able to retain high on-target activity on all sequences targetable by WT SpCas9, is not feasible. Kleinstiver et al. suggested that SpCas9-HF1 cleaves only "atypical", repetitive or homopolymeric targets with substantial off-target propensity[104]. According to our experiments, some typical targets that are not homopolymeric or repetitive sequences are also cleaved with high off-target effect by SpCas9-HF1 and some other increased fidelity variants as well. Our results clearly demonstrate that the only way to minimize off-target effects when it is required in case of any kind of target sites in any ranks (i.e. from *low*- to the *high-contributing*), is to have a full range of *lower-to-higher* increased fidelity SpCas9 variants in our toolbar.

Our results further demonstrated how using 5' G extended sgRNAs with WT SpCas9 could be an alternative solution for increasing the fidelity. Given the fact that extending the sgRNA with a 5' G nucleotide only modestly decreases the activity of WT SpCas9, on the basis of the interpretations provided here, we predict a considerably decreased off-target cleavage in case of *low-contributing* target sites. Furthermore, it is plausible to speculate that in contrast to the truncated sgRNAs[71], a matching G-extension may not generate new off-target sites compared to the unmodified sgRNA-WT SpCas9 complex (Fig. 11D, App. Fig. 2).

Another important conclusion of our work is that it is possible to further increase the fidelity of editing in case of increased fidelity variants by combining them with approaches that employ different mechanisms appropriate to increase fidelity, specifically by employing the SpCas9 variants in RNP form[96]. Demonstrating this principle could be important in case of applications where the reduction of the possibilities for incidental off-target events is of paramount priority such as in case of clinical applications.

Surprisingly, contrary to the claims made in the publication by Vakulskas et al.[120], all increased fidelity variants examined here showed similarly high activities with 20G-sgRNAs as the WT protein in RNP form (Fig. 14B - left boxplot). The Blackjack and *plus* variants generated by us cleaved the target sites in RNP form with 20G-sgRNAs the same level as their parental variants. Additionally, they were able to work with 21G-sgRNAs as well when delivered in RNP form. One of the possible explanations for the apparent discrepancy between our and Vakulskas's results is that they may have employed target sequences from the *low-contributing* target rank which are effectively cleaved only by the lower fidelity HiFi SpCas9 and WT SpCas9 but not by the middle/higher fidelity variants (i.e. e- and SpCas9-HF1) when delivered either as a plasmid or as RNP form. In this respect our results are highly significant showing that not only the lowest increased fidelity variants (WT, Blackjack-WT, Sniper[121] and HiFi SpCas9[120]) maintain their activity when delivered in the RNP form, but also higher increased fidelity variants do (i.e. e-, -HF1, B-evo-SpCas9 and the *plus* variants). Our data suggest that alternative fidelity-increasing methods (which employ different mechanism), such as the nickase approach could be combined with increased fidelity variants as well to reach even higher fidelity. However, since the binding activity of these increased fidelity nucleases does not seem to vary compared to WT (Fig. 7, 8), a combination with a dSpCas9-*Fok*I approach is less likely to cause more specific cleavage.

Blackjack and *plus* variants offer special advantages in all applications where a shortage of appropriate targets is a general problem. There is little room to maneuver, for example, when a specific position needs to be targeted by exploiting single strand oligos, when using either dCas9-*Fok*I nucleases or base editors or when tagging proteins. If only 20G-N19-NGG target sequences could be used that would lower the available target site to ¼ in such cases. Although some methods have been adapted to overcome that limitation[115,145,148,149], there is no general approach to extend the target space available for increased fidelity SpCas9 variants beyond the 20G-N19-NGG target sequences.

One of the most advantageous applications of the increased fidelity *plus* variants - for which currently no alternative exists -, is the usage of pooled sgRNA libraries to decrease false positive hits that frequently plague CRISPR screens caused by the off-target effect of WT SpCas9. There are no knockout pooled sgRNA libraries on Addgene.org available until these days that is restricted to only 20G-N19-NGG target

sequences, indeed many libraries are made up or contains 21G-sgRNAs, therefore they are compatible only with Blackjack or *plus* variants but not with their parental ones.

Incorporating Blackjack mutations into PAM-altered variants such as SpCas9-NG[164] and xCas9[93] may also be particularly beneficial to expand the available target space. These variants were designed to alter the constraint of the longer, NGG PAM motif required by SpCas9 to a more permissive NG PAM. However, the mutations incorporated to achieve this purpose has been shown to reduce the activity of xCas9 when used with 21G-sgRNAs (and increase its fidelity when used either with 20G- or 21G-sgRNAs)[121], limiting its usefulness. The available target sequences are particularly limited when base editors are used to modify nucleotides at specific positions. Combining Blackjack mutations either with xCas9- or SpCas9-NG-base editors would offer a much wider palette of targetable sites as the already existing ones.

Our findings are critical for the use of SpCas9-based technologies when maximal efficiency and fidelity are required. In case of the therapeutic usage of the SpCas9-based technologies it is of paramount priority to reduce the possibilities for incidental off-targets to a minimum, even below the detection limit of the currently existing approaches (<0.1% imposed by the current NGS technology). These applications frequently involve the optimization of a procedure based on exploiting one or a few targets that are at appropriate positions for their later routine use. The rankings observed here among these nucleases and among targets, and the extremely high fidelity of these increased fidelity nucleases on certain targets and the presentation how to increase their fidelity even further suggest that by careful designing, the off-target cleavage may be reduced even further to the level of incidentally-occurring off-target events.

6. Summary

We directly compared the already published increased fidelity nucleases (eSpCas9, SpCas9-HF1, HypaSpCas9, evoSpCas9) and the ones developed by us (Hypa2SpCas9, HeFSpCas9) in the same systems to understand the factors that affect their cleavage.

1. We found that the increased fidelity nucleases can be ranked by their fidelity that increases in parallel with their target-selectivity (i.e. the variant does not cleave or cleaves with reduced activity the target sites that are cleaved by the WT SpCas9 and that results in decreasing average activity and shrinking target space) as well, in the following order: WT SpCas9 < eSpCas9 < SpCas9-HF1 < HypaSpCas9 < Hypa2SpCas9 < evoSpCas9 < HeFSpCa9.

Our results suggest that not only the nucleases, but the target sequences can be ranked. At one end of the ranking there are *high-contributing* sequences that all variants can cleave, but only the highest fidelity variant cleaves them with minimal off-target cleavage. At the other end of the ranking there are *low-contributing* sequences that are cleaved efficiently only by the lower increased fidelity variants and by the WT SpCas9 but also with significantly lower level of off-target activity.

2. These rankings identified by us makes efforts to develop a superior nuclease variant that cleaves all target sites with high on-target activity and high specificity futile. Instead, our results show that it is necessary to have a series of variants with increasing fidelity in order to have an optimal nuclease for targets in all cleavability ranks.

3. We showed that increased fidelity variants can routinely only be used with sgRNAs containing perfectly matching 20 nucleotide-long spacers (20G-sgRNAs). 5' modified sgRNAs diminished the activities of increased fidelity variants.

4. We developed the **Blackjack** and *plus* increased fidelity variants which can be used routinely not only with 20G- but with 21G-sgRNAs as well. The *plus* variants show similar target-selectivity and fidelity pattern as their parental increased fidelity variants.

5. We showed that by using the RNP form of the increased fidelity variants instead of the plasmid form, their specificity can be further increased.

6. Acknowledgement

First and foremost, I would like to express my sincere gratitude to my supervisor Dr. Ervin Welker DSc for his continuous support over my masters and PhD years, which I spent in his research group. I would also like to thank him for his patience and the guidance which helped me in research and to acquire scientific thinking.

Special thanks to András Tálas, who was not only there when the project needed it but as a friend. He was always available for a good brainstorming about any scientific topic including our research or just chatting about Life, the Universe and Everything.

I am extremely thankful to Zsófia Rakvács, my wife who not just supported me but understood why I had to stay in the lab until midnight many times and helped me in experiments when I was overloaded.

I am grateful for the help what I got from Dr Elfrieda Fodor, Dr. Eszter Tóth, Krisztina Huszár, Zoltán Ligeti, Zsombor Welker, Nóra Weinhardt in my first author publications.

I am grateful to Ildikó Szűcsné Pulinka, Judit Szűcs, Bernadett Czene, Luca Pájer-Turgyán, Dávid Fetter for their excellent laboratory assistance. Without their help this work would have been much shorter.

I am thankful to Dr. György Váradi, Dr Edit Szabó and all the members of the 'FACS lab' for their valuable help and support.

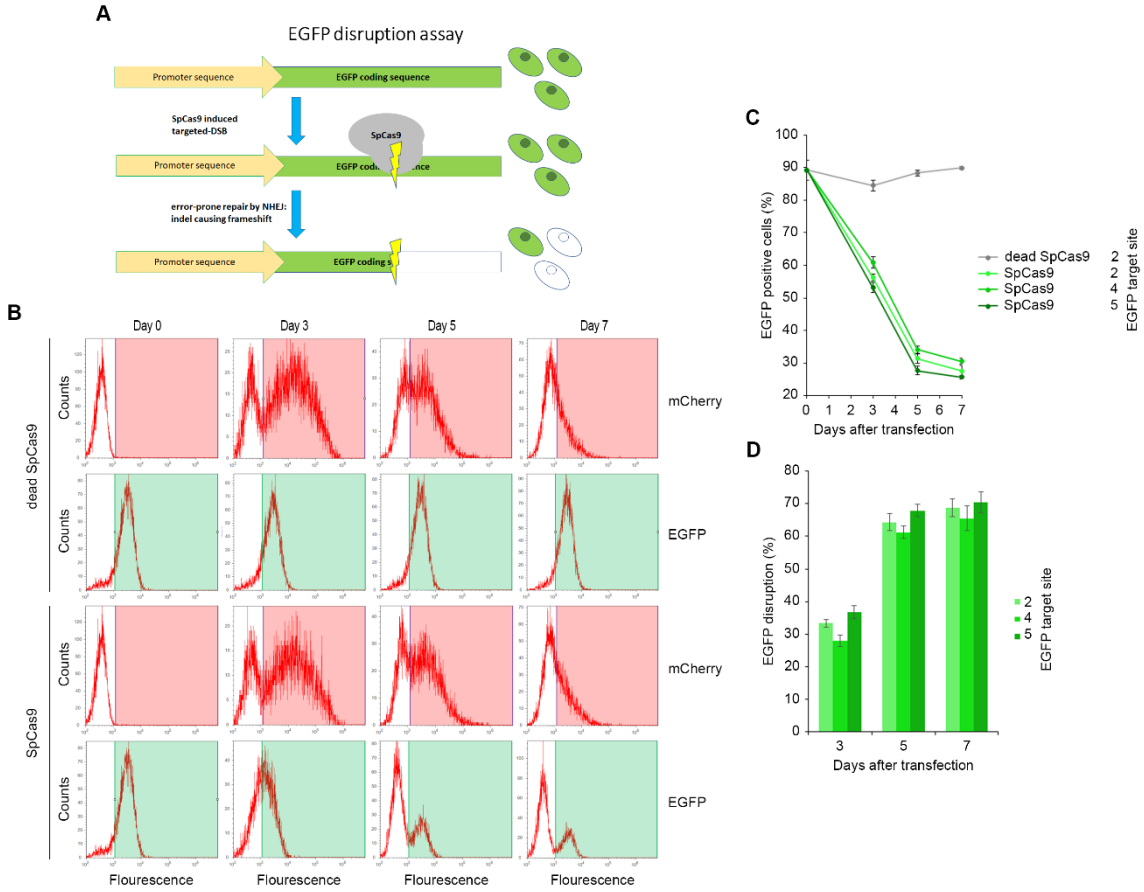
Special thanks to Dr. Ervin Welker, Dr. Eszter Tóth, Wes Lindinger, my father and my mother for the critics and corrections they provided while reading this thesis.

I wish to express many appreciations also to all my colleagues in the Institute of Enzymology, whom I met and worked with during my PhD years, especially to the people in the 'Szokásos péntek esti mulatozás' group, without them these years had been less fun. Special thanks to all my friends who understand my strange goal that I want to be a scientist.

And last but not least, I would like to thank my family, whose stood beside me, and supported me spiritually and financially as well to accomplish the present thesis.

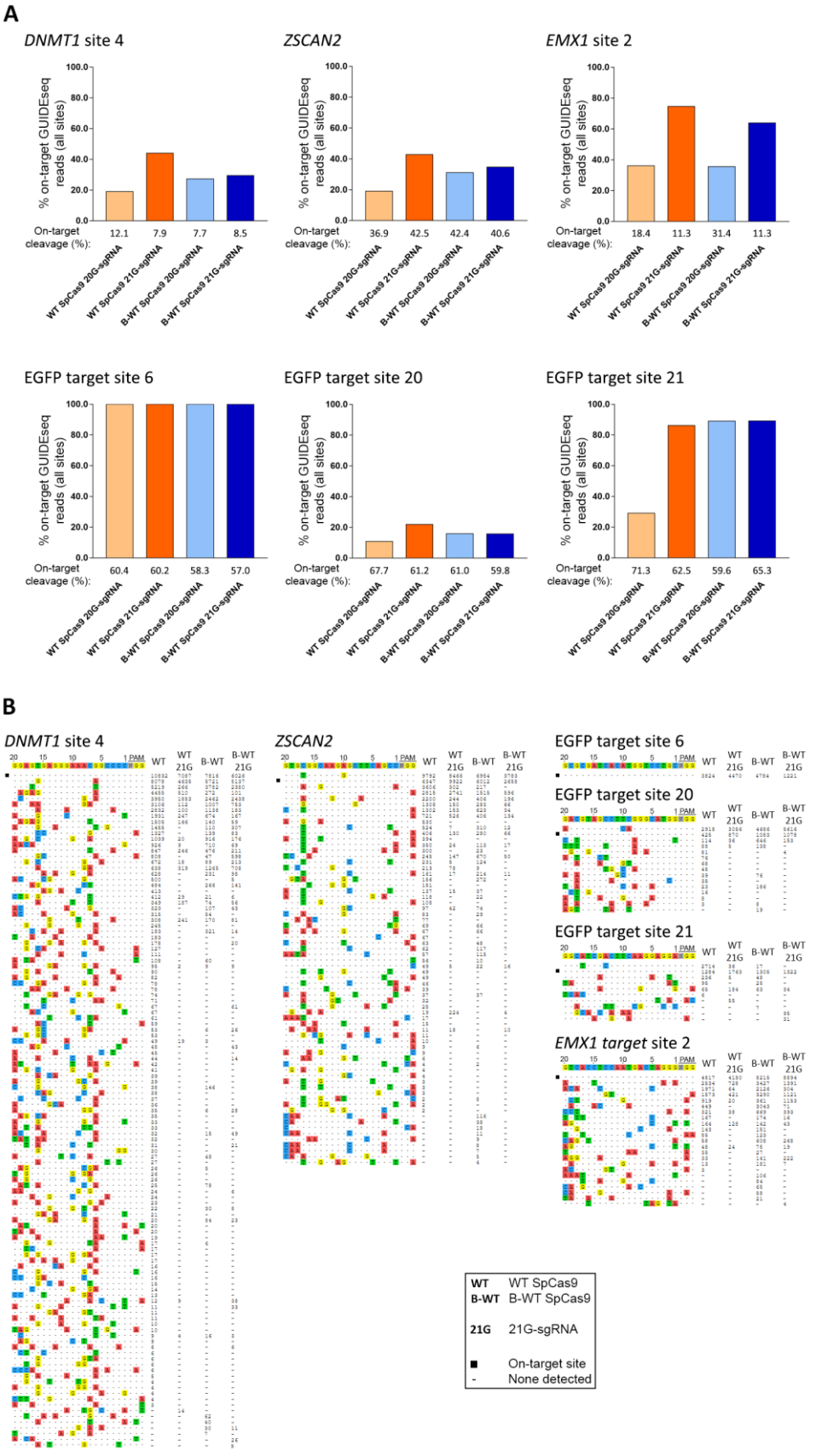
7. Appendix

Appendix Figure 1



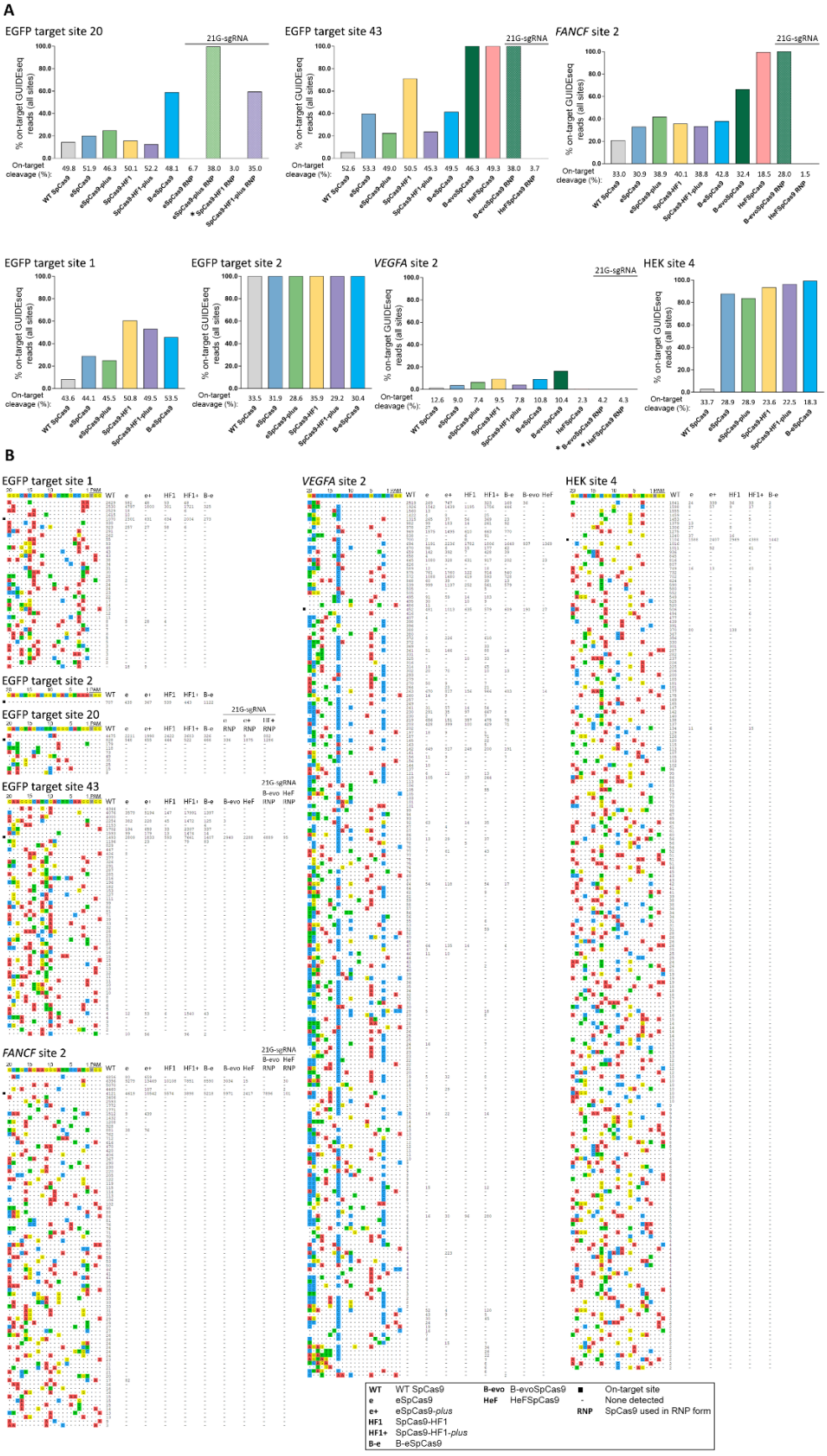
Appendix Figure 1. EGFP disruption induced by SpCas9 nuclease. A, Schematic overview of the EGFP disruption assay. After the SpCas9 induced double-strand break in a single integrated *EGFP* reporter coding sequence the error-prone NHEJ-mediated repair could lead to frameshift mutations that could disrupt the coding sequence and cause loss of fluorescence signal in cells. **B**, EGFP and mCherry FACS histograms of N2a.EGFP reporter cells co-transfected with mCherry/sgRNA expressing plasmid and either dead (targeted to EGFP site 2) or active SpCas9s (targeted to EGFP site 2) expressing plasmids, analyzed on days 0, 3, 5 and 7 post-transfection by flow cytometry. **C**, Percentages of EGFP positive cells at various days post-transfection derived from FACS analysis of samples transfected by either dead SpCas9 targeting site 2, or active SpCas9s targeting EGFP sites: 2, 4 or 5. **D**, EGFP disruption activities of wild type SpCas9s (targeted to EGFP sites 2, 4 or 5) measured on day 3, 5 and 7 post-transfection and calculated as described in Material and methods section.

Appendix Figure 2



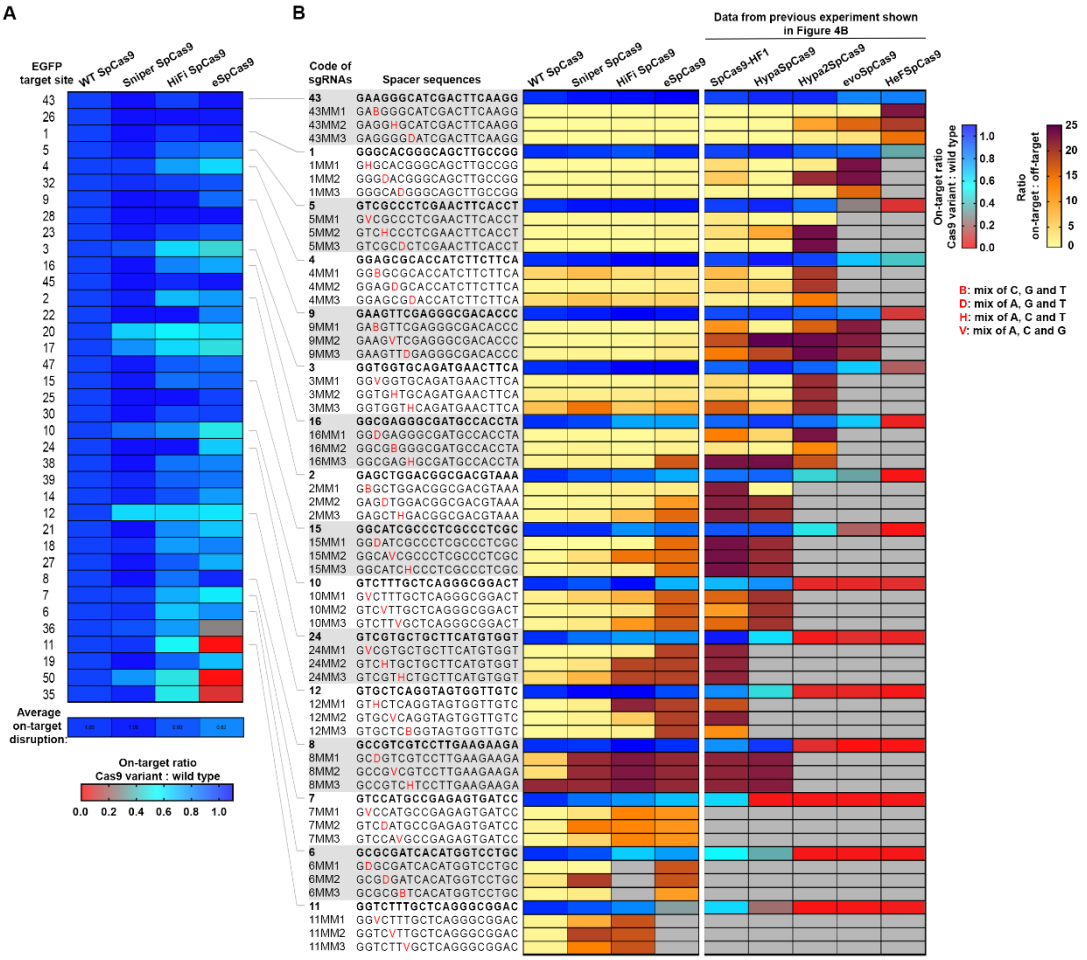
Appendix Figure 2. Both 5' G extended sgRNAs and Blackjack mutations increase the fidelity of WT SpCas9 as assessed by GUIDE-seq. Off-target cleavage sites of SpCas9 variants (Blackjack-WT and WT SpCas9) targeted either with 20G- or 21G-sgRNAs identified by GUIDE-seq. **A**, Specificity presented as the percentages of on-target reads per all reads captured by GUIDE-seq with the given sgRNAs. On-target cleavage activities were measured either by TIDE *(DNMT1* site 4, *ZSCAN2*, *EMX1* site 2) or flow cytometry (EGFP target site 6, 20 and 21) and are shown under the column charts. **B,** Read counts that give a measure of cleavage frequency at a given sequence shown; mismatched positions within the spacer or PAM are highlighted in color. (-) indicates zero reads, which means that off-target cleavage was not detected in the corresponding sample; black square indicates the on-target site.

Appendix Figure 3



Appendix Figure 3. The *plus* SpCas9 variants exhibit fidelity identical to their respective non-Blackjack nuclease variants eSpCas9 and SpCas9-HF1, as assessed by GUIDE-seq. Off-target cleavage sites of SpCas9 variants identified by GUIDE-seq. Seven sgRNAs targeted to either endogenous human genes or EGFP target sites. In case of RNP form 21G-sgRNAs were used, therefore the on-target activities of the non-Blackjack variants decrease. **A**, Specificity presented as the percentages of on-target reads per all reads captured by GUIDE-seq with the given sgRNAs. On-target cleavage activities were measured either by TIDE (*FANCF* site 2, *VEGFA* site 2, HEK site 4) or flow cytometry (EGFP target site 1, 2, 20 and 43) and are shown under the column charts. **B,** Read counts that give a measure of cleavage frequency at a given sequence shown; mismatched positions within the spacer or PAM are highlighted in color. (-) indicates zero reads, which means that off-target cleavage was not detected in the corresponding sample; black squares indicate on-target sites; data with no read counts are not shown (EGFP site 20: HF1 RNP; *VEGFA* site 2: B-evo RNP and HeF RNP).

Appendix Figure 4



Appendix Figure 4. Characterization of Sniper- and HiFi-SpCas9 compared to WT and eSpCas9. EGFP disruption activities of WT and increased fidelity nucleases. Sites are shown in the order of decreasing cleavability as in Figure 4A and B. **A,** Heatmap showing the relative on-target activities (red to blue) of the nuclease variants normalized to that of the wild type for each of the targets. The average on-target activities normalized to the wild type are shown under the map. **B**, Mismatch screen of the nuclease variants programmed with either perfectly matching (red to blue) or partially-mismatching 20G-sgRNAs [yellow to brown; a mixture of three different sgRNAs with the mismatch in same position[138]] as indicated in the figure. On *high-contributing* targets, Sniper, HiFi and eSpCas9 show similar specificities, all values being comparable to that of the WT protein. On *low-contributing* targets eSpCas9 shows higher specificity, while Sniper exhibits only marginally increased specificity compared to WT SpCas9. The specificity of HiFi SpCas9 is in-between that of Sniper and eSpCas9. The second panel of the heatmap derives from Figure 4B shown here to demonstrate where Sniper and HiFi SpCas9 fit into the rank.

Appendix: Table 1

	Name	Spacer length (nt)	Spacer Sequence	Target sequence with PAM
20th position G	EGFP site 1	20	GGGCACGGGCAGCTTGCCGG	GGGCACGGGCAGCTTGCCGGtgg
	EGFP site 2	20	GAGCTGGACGGCGACGTAAA	GAGCTGGACGGCGACGTAAAcgg
	EGFP site 3	20	GGTGGTGCAGATGAACTTCA	GGTGGTGCAGATGAACTTCAggg
	EGFP site 4	20	GGAGCGCACCATCTTCTTCA	GGAGCGCACCATCTTCTTCAagg
	EGFP site 5	20	GTCGCCCTCGAACTTCACCT	GTCGCCCTCGAACTTCACCTcgg
	EGFP site 6	20	GCGCGATCACATGGTCCTGC	GCGCGATCACATGGTCCTGCtgg
	EGFP site 7	20	GTCCATGCCGAGAGTGATCC	GTCCATGCCGAGAGTGATCCcgg
	EGFP site 8	20	GCCGTCGTCCTTGAAGAAGA	GCCGTCGTCCTTGAAGAAGAtgg
	EGFP site 9	20	GAAGTTCGAGGGCGACACCC	GAAGTTCGAGGGCGACACCCtgg
	EGFP site 10	20	GTCTTTGCTCAGGGCGGACT	GTCTTTGCTCAGGGCGGACTggg
	EGFP site 11	20	GGTCTTTGCTCAGGGCGGAC	GGTCTTTGCTCAGGGCGGACtgg
	EGFP site 12	20	GTGCTCAGGTAGTGGTTGTC	GTGCTCAGGTAGTGGTTGTCggg
	EGFP site 13	20	GCTGAAGGGCATCGACTTCA	GCTGAAGGGCATCGACTTCAagg
	EGFP site 14	20	GTACCAGCACTAGCCTCCTG	GTACCAGCACTAGCCTCCTGagg
	EGFP site 15	20	GGCATCGCCCTCGCCCTCGC	GGCATCGCCCTCGCCCTCGCcgg
	EGFP site 16	20	GGCGAGGGCGATGCCACCTA	GGCGAGGGCGATGCCACCTAcgg
	EGFP site 17	20	GCTGAAGCACTGCACGCCGT	GCTGAAGCACTGCACGCCGTagg
	EGFP site 18	20	GTGAACCGCATCGAGCTGAA	GTGAACCGCATCGAGCTGAAggg
	EGFP site 19	20	GTCAGGGTGGTCACGAGGGT	GTCAGGGTGGTCACGAGGGTggg
	EGFP site 20	20	GACGTAGCCTTCGGGCATGG	GACGTAGCCTTCGGGCATGGcgg
	EGFP site 21	20	GGCATCGACTTCAAGGAGGA	GGCATCGACTTCAAGGAGGAcgg
	EGFP site 22	20	GGTGTTCTGCTGGTAGTGGT	GGTGTTCTGCTGGTAGTGGTcgg
	EGFP site 23	20	GTGGTCACGAGGGTGGGCCA	GTGGTCACGAGGGTGGGCCAggg
	EGFP site 24	20	GTCGTGCTGCTTCATGTGGT	GTCGTGCTGCTTCATGTGGTcgg
	EGFP site 25	20	GGTCCGAGAGTCTGTAGCCA	GGTCCGAGAGTCTGTAGCCAtgg
	EGFP site 26	20	GCTGACGGTCAGGAGCCAGG	GCTGACGGTCAGGAGCCAGGagg
	EGFP site 27	20	GGCAGAGCAGGCTGACGGTC	GGCAGAGCAGGCTGACGGTCagg
	EGFP site 28	20	GAGCAGGCAGAGCAGGCTGA	GAGCAGGCAGAGCAGGCTGAcgg
	EGFP site 29	20	GAGGCCAGAGCAGGCAGAGC	GAGGCCAGAGCAGGCAGAGCagg
	EGFP site 30	20	GCTCTGCCTGCTCTGGCCTC	GCTCTGCCTGCTCTGGCCTCagg
	EGFP site 31	20	GGGCGAGGAGCTGTTCACCG	GGGCGAGGAGCTGTTCACCGggg
	EGFP site 32	20	GACCAGGATGGGCACCACCC	GACCAGGATGGGCACCACCCcgg
	EGFP site 33	20	GGTGCCCATCCTGGTCGAGC	GGTGCCCATCCTGGTCGAGCtgg
	EGFP site 34	20	GCCGTCCAGCTCGACCAGGA	GCCGTCCAGCTCGACCAGGAtgg
	EGFP site 35	20	GGCCACAAGTTCAGCGTGTC	GGCCACAAGTTCAGCGTGTCcgg
	EGFP site 36	20	GGTGGTCACGAGGGTGGGCC	GGTGGTCACGAGGGTGGGCCagg
	EGFP site 37	20	GGTCAGGGTGGTCACGAGGG	GGTCAGGGTGGTCACGAGGGtgg
	EGFP site 38	20	GTAGGTCAGGGTGGTCACGA	GTAGGTCAGGGTGGTCACGAggg
	EGFP site 39	20	GCACTGCACGCCGTAGGTCA	GCACTGCACGCCGTAGGTCAggg

	EGFP site 40	20	GCTTCATGTGGTCGGGGTAG	GCTTCATGTGGTCGGGGTAGcgg
	EGFP site 41	20	GCGCTCCTGGACGTAGCCTT	GCGCTCCTGGACGTAGCCTTcgg
	EGFP site 42	20	GGTGAACCGCATCGAGCTGA	GGTGAACCGCATCGAGCTGAagg
	EGFP site 43	20	GAAGGGCATCGACTTCAAGG	GAAGGGCATCGACTTCAAGGagg
	EGFP site 44	20	GATGCCGTTCTTCTGCTTGT	GATGCCGTTCTTCTGCTTGTcgg
	EGFP site 45	20	GCACGGGGCCGTCGCCGATG	GCACGGGGCCGTCGCCGATGggg
	EGFP site 46	20	GGTTGTCGGGCAGCAGCACG	GGTTGTCGGGCAGCAGCACGggg
	EGFP site 47	20	GTGGTTGTCGGGCAGCAGCA	GTGGTTGTCGGGCAGCAGCAcgg
	EGFP site 48	20	GGTGCTCAGGTAGTGGTTGT	GGTGCTCAGGTAGTGGTTGTcgg
	EGFP site 49	20	GTTGGGGTCTTTGCTCAGGG	GTTGGGGTCTTTGCTCAGGGcgg
	EGFP site 50	20	GCCGAGAGTGATCCCGGCGG	GCCGAGAGTGATCCCGGCGGcgg
20th position not G	EGFP site 51-no 5' G	20	CGATGCCCTTCAGCTCGATG	CGATGCCCTTCAGCTCGATGcgg
	EGFP site 52-no 5' G	20	CAACATCCTGGGGCACAAGC	CAACATCCTGGGGCACAAGCtgg
	EGFP site 53-no 5' G	20	CAGCTCGATGCGGTTCACCA	CAGCTCGATGCGGTTCACCAggg
	EGFP site 54-no 5' G	20	CAAGGAGGACGGCAACATCC	CAAGGAGGACGGCAACATCCtgg
	EGFP site 55-no 5' G	20	TCAGCTCGATGCGGTTCACC	TCAGCTCGATGCGGTTCACCagg
	EGFP site 56-no 5' G	20	AAGGAGGACGGCAACATCCT	AAGGAGGACGGCAACATCCTggg
	EGFP site 57-no 5' G	20	AGGAGGACGGCAACATCCTG	AGGAGGACGGCAACATCCTGggg
	EGFP site 58-no 5' G	20	CATGCCCGAAGGCTACGTCC	CATGCCCGAAGGCTACGTCCagg
	EGFP site 59-no 5' G	20	CGTGCTGCTTCATGTGGTCG	CGTGCTGCTTCATGTGGTCGggg
Series B: 21st position G mismatches	EGFP site 1B	21	gGGGCACGGGCAGCTTGCCGG	GGGCACGGGCAGCTTGCCGGtgg
	EGFP site 2B	21	gGAGCTGGACGGCGACGTAAA	GAGCTGGACGGCGACGTAAAcgg
	EGFP site 3B	21	gGGTGGTGCAGATGAACTTCA	GGTGGTGCAGATGAACTTCAggg
	EGFP site 4B	21	gGGAGCGCACCATCTTCTTCA	GGAGCGCACCATCTTCTTCAagg
	EGFP site 5B	21	gGTCGCCCTCGAACTTCACCT	GTCGCCCTCGAACTTCACCTcgg
	EGFP site 6B	21	gGCGCGATCACATGGTCCTGC	GCGCGATCACATGGTCCTGCtgg
	EGFP site 7B	21	gGTCCATGCCGAGAGTGATCC	GTCCATGCCGAGAGTGATCCcgg
	EGFP site 8B	21	gGCCGTCGTCCTTGAAGAAGA	GCCGTCGTCCTTGAAGAAGAtgg
	EGFP site 9B	21	gGAAGTTCGAGGGCGACACCC	GAAGTTCGAGGGCGACACCCtgg
	EGFP site 13B	21	gGCTGAAGGGCATCGACTTCA	GCTGAAGGGCATCGACTTCAagg
	EGFP site 14B	21	gGTACCAGCACTAGCCTCCTG	GTACCAGCACTAGCCTCCTGagg
	EGFP site 15B	21	gGGCATCGCCCTCGCCCTCGC	GGCATCGCCCTCGCCCTCGCcgg
	EGFP site 24B	21	gGTCGTGCTGCTTCATGTGGT	GTCGTGCTGCTTCATGTGGTcgg
	EGFP site 25B	21	gGGTCCGAGAGTCTGTAGCCA	GGTCCGAGAGTCTGTAGCCAtgg
	EGFP site 27B	21	gGGCAGAGCAGGCTGACGGTC	GGCAGAGCAGGCTGACGGTCagg
	EGFP site 28B	21	gGAGCAGGCAGAGCAGGCTGA	GAGCAGGCAGAGCAGGCTGAcgg
	EGFP site 29B	21	gGAGGCCAGAGCAGGCAGAGC	GAGGCCAGAGCAGGCAGAGCagg
	EGFP site 30B	21	gGCTCTGCCTGCTCTGGCCTC	GCTCTGCCTGCTCTGGCCTCagg
	EGFP site 31B	21	gGGGCGAGGAGCTGTTCACCG	GGGCGAGGAGCTGTTCACCGggg

	EGFP site 32B	21	gGACCAGGATGGGCACCACCC	GACCAGGATGGGCACCACCCcgg
	EGFP site 33B	21	gGGTGCCCATCCTGGTCGAGC	GGTGCCCATCCTGGTCGAGCtgg
	EGFP site 34B	21	gGCCGTCCAGCTCGACCAGGA	GCCGTCCAGCTCGACCAGGAtgg
	EGFP site 35B	21	gGGCCACAAGTTCAGCGTGTC	GGCCACAAGTTCAGCGTGTCcgg
	EGFP site 37B	21	gGGTCAGGGTGGTCACGAGGG	GGTCAGGGTGGTCACGAGGGtgg
	EGFP site 38B	21	gGTAGGTCAGGGTGGTCACGA	GTAGGTCAGGGTGGTCACGAggg
	EGFP site 39B	21	gGCACTGCACGCCGTAGGTCA	GCACTGCACGCCGTAGGTCAggg
	EGFP site 40B	21	gGCTTCATGTGGTCGGGGTAG	GCTTCATGTGGTCGGGGTAGcgg
	EGFP site 41B	21	gGCGCTCCTGGACGTAGCCTT	GCGCTCCTGGACGTAGCCTTcgg
	EGFP site 42B	21	gGGTGAACCGCATCGAGCTGA	GGTGAACCGCATCGAGCTGAagg
	EGFP site 43B	21	gGAAGGGCATCGACTTCAAGG	GAAGGGCATCGACTTCAAGGagg
	EGFP site 44B	21	gGATGCCGTTCTTCTGCTTGT	GATGCCGTTCTTCTGCTTGTcgg
	EGFP site 45B	21	gGCACGGGGCCGTCGCCGATG	GCACGGGGCCGTCGCCGATGggg
	EGFP site 46B	21	gGGTTGTCGGGCAGCAGCACG	GGTTGTCGGGCAGCAGCACGggg
	EGFP site 47B	21	gGTGGTTGTCGGGCAGCAGCA	GTGGTTGTCGGGCAGCAGCAcgg
	EGFP site 49B	21	gGTTGGGGTCTTTGCTCAGGG	GTTGGGGTCTTTGCTCAGGGcgg
	EGFP site 50B	21	gGCCGAGAGTGATCCCGGCGG	GCCGAGAGTGATCCCGGCGGcgg
Series C: 21st position G matches	EGFP site 10C	21	gGTCTTTGCTCAGGGCGGACT	gGTCTTTGCTCAGGGCGGACTggg
	EGFP site 11C	21	gGGTCTTTGCTCAGGGCGGAC	GGTCTTTGCTCAGGGCGGACtgg
	EGFP site 12C	21	gGTGCTCAGGTAGTGGTTGTC	gGTGCTCAGGTAGTGGTTGTCggg
	EGFP site 16C	21	gGGCGAGGGCGATGCCACCTA	gGGCGAGGGCGATGCCACCTAcgg
	EGFP site 17C	21	gGCTGAAGCACTGCACGCCGT	gGCTGAAGCACTGCACGCCGTagg
	EGFP site 18C	21	gGTGAACCGCATCGAGCTGAA	gGTGAACCGCATCGAGCTGAAggg
	EGFP site 19C	21	gGTCAGGGTGGTCACGAGGGT	gGTCAGGGTGGTCACGAGGGTggg
	EGFP site 20C	21	gGACGTAGCCTTCGGGCATGG	gGACGTAGCCTTCGGGCATGGcgg
	EGFP site 21C	21	gGGCATCGACTTCAAGGAGGA	gGGCATCGACTTCAAGGAGGAcgg
	EGFP site 22C	21	gGGTGTTCTGCTGGTAGTGGT	gGGTGTTCTGCTGGTAGTGGTcgg
	EGFP site 23C	21	gGTGGTCACGAGGGTGGGCCA	gGTGGTCACGAGGGTGGGCCAggg
	EGFP site 26C	21	gGCTGACGGTCAGGAGCCAGG	GCTGACGGTCAGGAGCCAGGagg
	EGFP site 36C	21	gGGTGGTCACGAGGGTGGGCC	GGTGGTCACGAGGGTGGGCCagg
	EGFP site 48C	21	gGGTGCTCAGGTAGTGGTTGT	GGTGCTCAGGTAGTGGTTGTcgg

Appendix Table 1. List of EGFP target sites and spacers. Modified sgRNAs targeting the identical EGFP sites, are named with the same number, but with an extension in the name (e.g. B, C, -no 5' G). Constructs cloned by me are yellow, blue color indicates constructs cloned by Krisztina Huszár.

Appendix: Table 2

	Name	Spacer length (nt)	Spacer Sequence	Sequence with PAM	References
	mouseSprn site 1B	21	gTTCTGCCCAGTAGGATGAAC	TTCTGCCCAGTAGGATGAACtgg	
	mouseSprn site 2B	21	gACTGGACTGCTGCCACGTGC	ACTGGACTGCTGCCACGTGCtgg	
	mouseSprn site 3B	21	gCTGGACTGCTGCCACGTGCT	CTGGACTGCTGCCACGTGCTggg	
	mouseSprn site 4B	21	gCACGTGCTGGGCTCTGCTGC	CACGTGCTGGGCTCTGCTGCtgg	
	mouseSprn site 5B	21	gCAGCAGCAGAGCCCAGCACG	CAGCAGCAGAGCCCAGCACGtgg	
	HEK site 4	20	GGCACTGCGGCTGGAGGTGG	GGCACTGCGGCTGGAGGTGGggg	
	VEGFA site 2	20	GACCCCCTCCACCCCGCCTC	GACCCCCTCCACCCCGCCTCcgg	from Casini et al., Nature Biotechnology 201
	FANCF site 2	20	GCTGCAGAAGGGATTCCATG	GCTGCAGAAGGGATTCCATGagg	from Kleinstiver et al., Nature 2016
	ZSCAN2	20	GTGCGGCAAGAGCTTCAGCC	GTGCGGCAAGAGCTTCAGCCggg	from Kleinstiver et al., Nature 2016
	ZSCAN2-21G	21	gGTGCGGCAAGAGCTTCAGCC	GTGCGGCAAGAGCTTCAGCCggg	from Kleinstiver et al., Nature 2016
	EMX1 site 2	20	GTCACCTCCAATGACTAGGG	GTCACCTCCAATGACTAGGGtgg	
	EMX1 site 2-21G	21	gGTCACCTCCAATGACTAGGG	GTCACCTCCAATGACTAGGGtgg	from Kleinstiver et al., Nature 2016
	DNMT1 site 4	20	GGAGTGAGGGAAACGGCCCC	GGAGTGAGGGAAACGGCCCCagg	
	DNMT1 site 4-21G	21	gGGAGTGAGGGAAACGGCCCC	GGAGTGAGGGAAACGGCCCCagg	from Chen et al., Nature 2017
sgRNAs used for upregulation (B: 21st, 15th position mismatched G, C: 21st position matched G	mousePRNPpromoter site 1C	21	gCGGGGCGTGATGCTCACCAA	gCGGGGCGTGATGCTCACCAAtgg	
	mousePRNPpromoter site 2B	21	gGTGGAAAGGGGCGGGGCAGG	GTGGAAAGGGGCGGGGCAGGcgg	
	mousePRNPpromoter site 3B	21	gGCATTTAAGCCAGTCCGGAG	GCATTTAAGCCAGTCCGGAGcgg	
	mousePRNPpromoter site 4B	21	gATAGTTGCTGAGCGTCGTCA	ATAGTTGCTGAGCGTCGTCAggg	
	mousePRNPpromoter site 5B	21	gCTCAACTACCCATTATGTAA	CTCAACTACCCATTATGTAAcgg	
	mousePRNPpromoter site 6B	15	gAGGGGCGGGGCAGG	AGGGGCGGGGCAGGcgg	
	mousePRNPpromoter site 7B	15	gGTGATGCTCACCAA	GTGATGCTCACCAAtgg	
	mousePRNPpromoter site 8B	15	gAAGCCAGTCCGGAG	AAGCCAGTCCGGAGcgg	
	mousePRNPpromoter site 9B	15	gGCTGAGCGTCGTCA	GCTGAGCGTCGTCAggg	
	mousePRNPpromoter site 10B	15	gTACCCATTATGTAA	TACCCATTATGTAAcgg	

Appendix Table 2. List of endogenous target sites and spacers. 5' modified sgRNAs have an extension in the name (e.g. B, C, 21G, same nomenclature as in App. Table 1). Constructs cloned by me are yellow, green color indicates constructs cloned by András Tálas.

Appendix: Table 3

On target site	Code of sgRNAs	Mixed mismatched spacers	Spacer length (nt)	Mix of spacer sequences
EGFP site 1	1MM1	1-G19H	20	GHGCACGGGCAGCTTGCCGG
	1MM2	1-C17D	20	GGGDACGGGCAGCTTGCCGG
	1MM3	1-C15D	20	GGGCADGGGCAGCTTGCCGG
EGFP site 2	2MM1	2-A19B	20	GBGCTGGACGGCGACGTAAA
	2MM2	2-C17D	20	GAGDTGGACGGCGACGTAAA
	2MM3	2-G15H	20	GAGCTHGACGGCGACGTAAA
EGFP site 3	3MM1	3-T18V	20	GGVGGTGCAGATGAACTTCA
	3MM2	3-G16H	20	GGTGHTGCAGATGAACTTCA
	3MM3	3-G14H	20	GGTGGTHCAGATGAACTTCA
EGFP site 4	4MM1	4-A18B	20	GGBGCGCACCATCTTCTTCA
	4MM2	4-C16D	20	GGAGDGCACCATCTTCTTCA
	4MM3	4-C14D	20	GGAGCGDACCATCTTCTTCA
EGFP site 5	5MM1	5-T19V	20	GVCGCCCTCGAACTTCACCT
	5MM2	5-G17H	20	GTCHCCCTCGAACTTCACCT
	5MM3	5-C15D	20	GTCGCDCTCGAACTTCACCT
EGFP site 6	6MM1	6-C19D	20	GDGCGATCACATGGTCCTGC

	6MM2	6-C17D	20	GCGDGATCACATGGTCCTGC
	6MM3	6-A15B	20	GCGCGBTCACATGGTCCTGC
EGFP site 7	7MM1	7-T19V	20	GVCCATGCCGAGAGTGATCC
	7MM2	7-C17D	20	GTCDATGCCGAGAGTGATCC
	7MM3	7-T15V	20	GTCCAVGCCGAGAGTGATCC
EGFP site 8	8MM1	8-C18D	20	GCDGTCGTCCTTGAAGAAGA
	8MM2	8-T16V	20	GCCGVCGTCCTTGAAGAAGA
	8MM3	8-G14H	20	GCCGTCHTCCTTGAAGAAGA
EGFP site 9	9MM1	9-A18B	20	GABGTTCGAGGGCGACACCC
	9MM2	9-T16V	20	GAAGVTCGAGGGCGACACCC
	9MM3	9-C14D	20	GAAGTTDGAGGGCGACACCC
EGFP site 10	10MM1	10-T19V	20	GVCTTTGCTCAGGGCGGACT
	10MM2	10-T17V	20	GTCVTTGCTCAGGGCGGACT
	10MM3	10-T15V	20	GTCTTVGCTCAGGGCGGACT
EGFP site 11	11MM1	11-T18V	20	GGVCTTTGCTCAGGGCGGAC
	11MM2	11-T16V	20	GGTCVTTGCTCAGGGCGGAC
	11MM3	11-T14V	20	GGTCTTVGCTCAGGGCGGAC
EGFP site 12	12MM1	12-G18H	20	GTHCTCAGGTAGTGGTTGTC
	12MM2	12-T16V	20	GTGCVCAGGTAGTGGTTGTC
	12MM3	12-A14B	20	GTGCTCBGGTAGTGGTTGTC
EGFP site 15	15MM1	15-C18D	20	GGDATCGCCCTCGCCCTCGC
	15MM2	15-T16V	20	GGCAVCGCCCTCGCCCTCGC
	15MM3	15-G14H	20	GGCATCHCCCTCGCCCTCGC
EGFP site 16	16MM1	16-C18D	20	GGDGAGGGCGATGCCACCTA
	16MM2	16-A16B	20	GGCGBGGGCGATGCCACCTA
	16MM3	16-G14H	20	GGCGAGHGCGATGCCACCTA
EGFP site 24	24MM1	24-T19V	20	GVCGTGCTGCTTCATGTGGT
	24MM2	24-G17H	20	GTCHTGCTGCTTCATGTGGT
	24MM3	24-G15H	20	GTCGTHCTGCTTCATGTGGT
EGFP site 43	43MM1	43-A18B	20	GABGGGCATCGACTTCAAGG
	43MM2	43-G16H	20	GAGGHGCATCGACTTCAAGG
	43MM3	43-C14D	20	GAGGGGDATCGACTTCAAGG

Appendix Table 3. List of mismatching sgRNAs. All sgRNAs contain a single mismatched 20 base long spacer sequence targeting the EGFP coding sequence. The name of the mixed mismatched spacers indicate the target site (e.g. 1 – EGFP target site 1), the position mismatched (e.g. 1-G19) and the possible mismatches (e.g. 1-G19H; **B**: mix of C, G and T; **D**: mix of A, G and T; **H**: mix of A, C and T; **V**: mix of A, C and G).

Appendix: Table 4

nuclease description	Sub #1	Sub #2	Sub #3	Sub #4	Sub #5	Sub #6	Sub #7	Blackjack (B) mutation/deletion/insertion	mammalian expression plasmids	bacterial expression plasmids
									Addgene ID	Addgene ID
dSpCas9	D10A	H840A						-	#92112	
WT SpCas9								-	#126753	#
eSpCas9	K848A	K1003A	R1060A					-	#126754	#126769
SpCas9-HF1	N497A	R661A	Q695A	Q926A				-	#126755	#126770
HypaSpCas9	N692A	M694A	Q695A	H698A				-	#126756	
Hypa2SpCas9	R661A	N692A	M694A	Q695A	H698A			-	#126757	
evoSpCas9	M495V	Y515N	K526E	R661Q				-	#126758	
HeFSpCas9	N497A	R661A	Q695A	K848A	Q926A	K1003A	R1060A	-	#126759	#126771
HF-B1	N497A	R661A	Q695A	Q926A				E1007G; Y1013G		
HF-B2	N497A	R661A	Q695A	Q926A				K1003-G-Δ LESEFVYGDY-K1014		
HF-B3	N497A	R661A	Q695A	Q926A				K1003-Δ LESEFVYGDYKVY-D1017		
HF-B4	N497A	R661A	Q695A	Q926A				K1003-G-Δ LESEFVYGDYKVY-D1017		
HF-B5	N497A	R661A	Q695A	Q926A				K1003-GK-Δ LESEFVYGDYKVY-D1017		
HF-B6	N497A	R661A	Q695A	Q926A				L1004-Δ ESEFVYGDY-K1014		
HF-B7	N497A	R661A	Q695A	Q926A				L1004-G-Δ ESEFVYGDY-K1014		
HF-B8	N497A	R661A	Q695A	Q926A				L1004-P-Δ ESEFVYGDY-K1014		
B-SpCas9-HF1 (HF-B9)	N497A	R661A	Q695A	Q926A				L1004-GG-Δ ESEFVYGDY-K1014	#126762	
HF-B10	N497A	R661A	Q695A	Q926A				L1004-GP-Δ ESEFVYGDY-K1014		
HF-B11	N497A	R661A	Q695A	Q926A				L1004-GGG-Δ ESEFVYGDY-K1014		
HF-B12	N497A	R661A	Q695A	Q926A				L1004-GGGG-Δ ESEFVYGDY-K1014		
HF-B13	N497A	R661A	Q695A	Q926A				L1004-GG-Δ ESEFVYGDYK-V1015		
HF-B14	N497A	R661A	Q695A	Q926A				L1004-GG-Δ ESEFVYGDYKV-Y1016		
HF-B15	N497A	R661A	Q695A	Q926A				L1004-GG-Δ ESEFVYGDYKVY-D1017		
HF-B16	N497A	R661A	Q695A	Q926A				L1004-GG-Δ EFVYGDY-K1014		
HF-B17	N497A	R661A	Q695A	Q926A				L1004-GG-Δ EFVYGDYKVY-D1017		
HF-B18	N497A	R661A	Q695A	Q926A				S1006-Δ EFVYGDY-K1014		
HF-B19	N497A	R661A	Q695A	Q926A				S1006-G-Δ EFVYGDY-K1014		
B-WT SpCas9								L1004-GG-Δ ESEFVYGDY-K1014	#126760	
B-eSpCas9	K848A	K1003A	R1060A					L1004-GG-Δ ESEFVYGDY-K1014	#126761	#126772
B-HypaSpCas9	N692A	M694A	Q695A	H698A				L1004-GG-Δ ESEFVYGDY-K1014	#126763	
B-Hypa2SpCas9	R661A	N692A	M694A	Q695A	H698A			L1004-GG-Δ ESEFVYGDY-K1014	#126764	
B-evoSpCas9	M495V	Y515N	K526E	R661Q				L1004-GG-Δ ESEFVYGDY-K1014	#126765	#126773
B-HeFSpCas9	N497A	R661A	Q695A	K848A	Q926A	K1003A	R1060A	L1004-GG-Δ ESEFVYGDY-K1014	#126766	
e+1		K1003A	R1060A					L1004-GG-Δ ESEFVYGDY-K1014		
eSpCas9-plus (e+2)	K848A		R1060A					L1004-GG-Δ ESEFVYGDY-K1014	#126767	#126774
e+3	K848A	K1003A						L1004-GG-Δ ESEFVYGDY-K1014		
e+4	K848A							L1004-GG-Δ ESEFVYGDY-K1014		
HF+1		R661A	Q695A	Q926A				L1004-GG-Δ ESEFVYGDY-K1014		
SpCas9-HF1-plus (HF+2)	N497A		Q695A	Q926A				L1004-GG-Δ ESEFVYGDY-K1014	#126768	#126775
HF+3	N497A	R661A		Q926A				L1004-GG-Δ ESEFVYGDY-K1014		
HF+4	N497A	R661A	Q695A					L1004-GG-Δ ESEFVYGDY-K1014		
HF+5	N497A			Q926A				L1004-GG-Δ ESEFVYGDY-K1014		
HF+6	N497A		Q695A					L1004-GG-Δ ESEFVYGDY-K1014		
HF+7	N497A	R661A						L1004-GG-Δ ESEFVYGDY-K1014		
Sniper SpCas9	F539S	M763I	K890N					-	#126777	
HiFi SpCas9	R691A							-	#126778	

Appendix Table 4. List of SpCas9 variants cloned and examined. Constructs cloned by me are yellow, green color indicates constructs cloned by András Tálas, blue color by Krisztina Huszár.

Appendix: Table 5

Name of plasmid construct	Addgene ID
pX330-U6-Chimeric_BB-CBh-hSpCas9	42230
eSpCas9(1.1)	71814
VP12	72247
pX335-U6-Chimeric_BB-CBh-hSpCas9n(D10A)	42335
pET-dCas9-VP64-6xHis	62935
sgRNA(MS2) cloning backbone	61424
MS2-P65-HSF1_GFP	61423
pMJ806	39312
pBMN DHFR(DD)-YFP	29325
pX330-Flag-dSpCas9	92113
pX330-Flag-deSpCas9	92114
pX330-Flag-dSpCas9-HF1	92115
pX330-Flag-dHeFSpCas9	92116
pmCherry_gRNA	80457
U6-sgRNA(MS2)_EF1a-MS2-P65-HSF1	92120
pET-dead eSpCas9-VP64-6xHis	92117
pET-dead SpCas9-HF1-VP64-6xHis	92118
pET-dead HeFSpCas9-VP64-6xHis	92119

Appendix Table 5. List of additional plasmids used (either ordered from Addgene.org or cloned by us). Constructs ordered from Addgene.org are yellow, green color indicates constructs cloned by us.

Appendix: Table 6

Name of primers	Sequence
T7-+G-Site51	AAAAAATAATACGACTCACTATAgCGATGCCCTTCAGCTCGATG
T7-+G-Site52	AAAAAATAATACGACTCACTATAgCAACATCCTGGGGCACAAGC
T7-+G-Site53	AAAAAATAATACGACTCACTATAgCAGCTCGATGCGGTTCACCA
T7-+G-Site54	AAAAAATAATACGACTCACTATAgCAAGGAGGACGGCAACATCC
T7-+G-Site55	AAAAAATAATACGACTCACTATAgTCAGCTCGATGCGGTTCACC
T7-+G-Site56	AAAAAATAATACGACTCACTATAgAAGGAGGACGGCAACATCCT
T7-+G-Site57	AAAAAATAATACGACTCACTATAgAGGAGGACGGCAACATCCTG
T7-+G-Site58	AAAAAATAATACGACTCACTATAgCATGCCCGAAGGCTACGTCC
T7-+G-Site59	AAAAAATAATACGACTCACTATAgCGTGCTGCTTCATGTGGTCG
T7-+2G-Site51	AAAAAATAATACGACTCACTATAggCGATGCCCTTCAGCTCGATG
T7-+2G-Site52	AAAAAATAATACGACTCACTATAggCAACATCCTGGGGCACAAGC
T7-+2G-Site53	AAAAAATAATACGACTCACTATAggCAGCTCGATGCGGTTCACCA
T7-+2G-Site54	AAAAAATAATACGACTCACTATAggCAAGGAGGACGGCAACATCC
T7-+2G-Site55	AAAAAATAATACGACTCACTATAggTCAGCTCGATGCGGTTCACC
T7-+2G-Site56	AAAAAATAATACGACTCACTATAggAAGGAGGACGGCAACATCCT
T7-+2G-Site57	AAAAAATAATACGACTCACTATAggAGGAGGACGGCAACATCCTG
T7-+2G-Site58	AAAAAATAATACGACTCACTATAggCATGCCCGAAGGCTACGTCC
T7-+2G-Site59	AAAAAATAATACGACTCACTATAggCGTGCTGCTTCATGTGGTCG
T7-Site-1	AAAAAATAATACGACTCACTATAgggcacgggcagcttgccgg
T7-Site3	AAAAAATAATACGACTCACTATAGGTGGTGCAGATGAACTTCA
T7-Site4	AAAAAATAATACGACTCACTATAggagcgcaccatcttcttca
T7-Site16	AAAAAATAATACGACTCACTATAggcgagggcgatgccaccta
T7-Site21	AAAAAATAATACGACTCACTATAGGCATCGACTTCAAGGAGGA
T7-Site22	AAAAAATAATACGACTCACTATAGGTGTTCTGCTGGTAGTGGT
T7-Site31	AAAAAATAATACGACTCACTATAGGGCGAGGAGCTGTTCACCG
T7-Site33	AAAAAATAATACGACTCACTATAGGTGCCCATCCTGGTCGAGC
T7-Site37	AAAAAATAATACGACTCACTATAGGTCAGGGTGGTCACGAGGG
T7-Site46	AAAAAATAATACGACTCACTATAGGTTGTCGGGCAGCAGCACG
T7-Site48	AAAAAATAATACGACTCACTATAGGTGCTCAGGTAGTGGTTGT
T7-+G-Site1	AAAAAATAATACGACTCACTATAgGGGCACGGGCAGCTTGCCGG
T7-+G-Site3	AAAAAATAATACGACTCACTATAgGGTGGTGCAGATGAACTTCA
T7-+G-Site4	AAAAAATAATACGACTCACTATAgGGAGCGCACCATCTTCTTCA
T7-+G-Site16	AAAAAATAATACGACTCACTATAgGGCGAGGGCGATGCCACCTA
T7-+G-Site21	AAAAAATAATACGACTCACTATAgGGCATCGACTTCAAGGAGGA
T7-+G-Site22	AAAAAATAATACGACTCACTATAgGGTGTTCTGCTGGTAGTGGT
T7-+G-Site31	AAAAAATAATACGACTCACTATAgGGGCGAGGAGCTGTTCACCG
T7-+G-Site33	AAAAAATAATACGACTCACTATAgGGTGCCCATCCTGGTCGAGC
T7-+G-Site37	AAAAAATAATACGACTCACTATAgGGTCAGGGTGGTCACGAGGG
T7-+G-Site46	AAAAAATAATACGACTCACTATAgGGTTGTCGGGCAGCAGCACG
T7-+G-Site48	AAAAAATAATACGACTCACTATAgGGTGCTCAGGTAGTGGTTGT

T7-Site5	AAAAAATAATACGACTCACTATAgtcgccctcgaacttcacct
T7-Site13	AAAAAATAATACGACTCACTATAGCTGAAGGGCATCGACTTCA
T7-Site17	AAAAAATAATACGACTCACTATAGCTGAAGCACTGCACGCCGT
T7-Site18	AAAAAATAATACGACTCACTATAGTGAACCGCATCGAGCTGAA
T7-Site20	AAAAAATAATACGACTCACTATAGACGTAGCCTTCGGGCATGG
T7-Site23	AAAAAATAATACGACTCACTATAGTGGTCACGAGGGTGGGCCA
T7-Site38	AAAAAATAATACGACTCACTATAGTAGGTCAGGGTGGTCACGA
T7-Site39	AAAAAATAATACGACTCACTATAGCACTGCACGCCGTAGGTCA
T7-Site44	AAAAAATAATACGACTCACTATAGATGCCGTTCTTCTGCTTGT
T7-Site45	AAAAAATAATACGACTCACTATAGCACGGGGCCGTCGCCGATG
T7-Site47	AAAAAATAATACGACTCACTATAGTGGTTGTCGGGCAGCAGCA
T7-+G-Site5	AAAAAATAATACGACTCACTATAGgtcgccctcgaacttcacct
T7-+G-Site13	AAAAAATAATACGACTCACTATAgGCTGAAGGGCATCGACTTCA
T7-+G-Site17	AAAAAATAATACGACTCACTATAgGCTGAAGCACTGCACGCCGT
T7-+G-Site18	AAAAAATAATACGACTCACTATAgGTGAACCGCATCGAGCTGAA
T7-+G-Site20	AAAAAATAATACGACTCACTATAgGACGTAGCCTTCGGGCATGG
T7-+G-Site23	AAAAAATAATACGACTCACTATAgGTGGTCACGAGGGTGGGCCA
T7-+G-Site38	AAAAAATAATACGACTCACTATAgGTAGGTCAGGGTGGTCACGA
T7-+G-Site39	AAAAAATAATACGACTCACTATAgGCACTGCACGCCGTAGGTCA
T7-+G-Site44	AAAAAATAATACGACTCACTATAgGATGCCGTTCTTCTGCTTGT
T7-+G-Site45	AAAAAATAATACGACTCACTATAgGCACGGGGCCGTCGCCGATG
T7-+G-Site47	AAAAAATAATACGACTCACTATAgGTGGTTGTCGGGCAGCAGCA
T7-+G-Site43	AAAAAATAATACGACTCACTATAgGAAGGGCATCGACTTCAAGG
T7-+G-VEGFAsite2	AAAAAATAATACGACTCACTATAgGACCCCCTCCACCCCGCCTC
T7-+G-FANCFsite2	AAAAAATAATACGACTCACTATAgGCTGCAGAAGGGATTCCATG
T7-forward-EGFP site 10C sgRNA	aaaaaaTAATACGACTCACTATAgGTCTTTGCTCAGGGCGGAC
T7-forward-EGFP site 12C sgRNA	aaaaaaTAATACGACTCACTATAgGTGCTCAGGTAGTGGTTGTCg
T7-forward-EGFP site 16 sgRNA	AAAAAATAATACGACTCACTATAggcgagggcgatgccac
T7-forward-EGFP site 21 sgRNA	AAAAAATAATACGACTCACTATAGGCATCGACTTCAAGGAGGAg
T7-forward-EGFP site 22 sgRNA	AAAAAATAATACGACTCACTATAGGTGTTCTGCTGGTAGTGGTg
T7-reverse primer	AAAAAAGCACCGACTCGGTGCC

Appendix Table 6. List of primers used to amplify sgRNA sequences for *in vitro* transcription

Appendix: Table 7

DNA Target	Name of primers	Sequence
EMSA-1 (EGFP target stie 10, 12, 22)	EMSA1-for	GTACAACTACAACAGCCACAACG
	EMSA1-rev	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGCTCCGGACTATAGCTCATCT
EMSA-2 (EGFP target stie 21)	EMSA2-for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCACATGAAGCAGCACGACT
	EMSA2-rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTCTTCTGCTTGTCGGCCAT
EMSA-3 (EGFP target stie 16)	EMSA2-for	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACGTAAACGGCCACAAGTTC
	EMSA2-rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCTTGTAGTTGCCGTCGTCC

Appendix Table 7. List of primers used to amplify target sequences for electrophoretic mobility shift assay (EMSA). PCR template sequence was an EGFP-KDEL coding sequence containing plasmid.

Appendix: Table 8

Name of primers	Sequence	References
FANCF site 2-for	GGGCCGGGAAAGAGTTGCTG	from Kleinstiver et al., Nature 2016
FANCF site 2-rev	GCCCTACATCTGCTCTCCCTCC	from Kleinstiver et al., Nature 2016
VEGFA site 2-for	CGAGGAAGAGAGAGACGGGGTC	from Chen et al., Nature 2017
VEGFA site 2-rev	CTCCAATGCACCCAAGACAGCAG	from Chen et al., Nature 2017
HEK site 4-for	AGAGAAGTTGGAGTGAAGGCAGAG	from Casini et al., Nature Biotechnology 2018
HEK site 4-rev	GTCAGACGTCCAAAACCAGACTCC	from Casini et al., Nature Biotechnology 2018
ZSCAN2-for	AGTGTGGGGTGTGTGGGAAG	from Kleinstiver et al., Nature 2016
ZSCAN2-rev	ACGGGACTTGACTCAGACCACT	from Kleinstiver et al., Nature 2016
EMX1 site 2-for	GGAGCAGCTGGTCAGAGGGG	from Kleinstiver et al., Nature 2016
EMX1 site 2-rev	CCATAGGGAAGGGGGACACTGG	from Kleinstiver et al., Nature 2016
DNMT1 site 4-for	CCAGAATGCACAAAGTACTGCAC	from Chen et al., Nature 2017
DNMT1 site 4-rev	GCCAAAGCCCGAGAGAGTGCC	from Chen et al., Nature 2017

Appendix Table 8. List of primers used to amplify sequence of interest for TIDE

Appendix: Table 9

Name	Target sequence with PAM	21G-sgRNAs (Spacer sequence)	Ribosyme flanked-sgRNAs (Ribosyme-Spacer sequence)	tRNA flanked-sgRNAs (Spacer sequence)
EGFP site 51B	CGATGCCCTTCAGCTCGATGcgg	gCGATGCCCTTCAGCTCGATG	GCATCGCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCCGATGCCCTTCAGCTCGATG	CGATGCCCTTCAGCTCGATG
EGFP site 52C	CAACATCCTGGGGCACAAGCtgg	gCAACATCCTGGGGCACAAGC	ATGTTGCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCCAACATCCTGGGGCACAAGC	CAACATCCTGGGGCACAAGC
EGFP site 53B	CAGCTCGATGCGGTTCACCAgggg	gCAGCTCGATGCGGTTCACCA	GAGCTGCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCCAGCTCGATGCGGTTCACCA	CAGCTCGATGCGGTTCACCA
EGFP site 59B	CGTGCTGCTTCATGTGGTCGgggg	gCGTGCTGCTTCATGTGGTCG	AGCACGCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCCGTGCTGCTTCATGTGGTCG	CGTGCTGCTTCATGTGGTCG
EGFP site 60C	CCGTCCAGCTCGACCAGGATggg	gCCGTCCAGCTCGACCAGGAT	GGACGGCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCCCGTCCAGCTCGACCAGGAT	CCGTCCAGCTCGACCAGGAT
EGFP site 61C	CAACTACAAGACCCGCGCCGagg	gCAACTACAAGACCCGCGCCG	TAGTTGCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCCAACTACAAGACCCGCGCCG	CAACTACAAGACCCGCGCCG
EGFP site 62B	AAGGGCGAGGAGCTGTTCACCgg	gAAGGGCGAGGAGCTGTTCAC	GCCCTTCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCAAGGGCGAGGAGCTGTTCAC	AAGGGCGAGGAGCTGTTCAC
EGFP site 63B	AAGTTCAGCGTGTCCGGCGAggg	gAAGTTCAGCGTGTCCGGCGA	GAACTTCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCAAGTTCAGCGTGTCCGGCGA	AAGTTCAGCGTGTCCGGCGA
EGFP site 64B	AGCGTGTCCGGCGAGGGCGAggg	gAGCGTGTCCGGCGAGGGCGA	CACGCTCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCAGCGTGTCCGGCGAGGGCGA	AGCGTGTCCGGCGAGGGCGA
EGFP site 65B	ACGAGGGTGGGCCAGGGCACggg	gACGAGGGTGGGCCAGGGCAC	CCTCGTCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCACGAGGGTGGGCCAGGGCAC	ACGAGGGTGGGCCAGGGCAC
EGFP site 66B	AGAAGTCGTGCTGCTTCATGtgg	gAGAAGTCGTGCTGCTTCATG	ACTTCTCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCAGAAGTCGTGCTGCTTCATG	AGAAGTCGTGCTGCTTCATG
EGFP site 67B	ACCATCTTCTTCAAGGACGAcgg	gACCATCTTCTTCAAGGACGA	GATGGTCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCACCATCTTCTTCAAGGACGA	ACCATCTTCTTCAAGGACGA
EGFP site 68B	AAGGAGGACGGCAACATCCTggg	gAAGGAGGACGGCAACATCCT	CTCCTTCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCAAGGAGGACGGCAACATCCT	AAGGAGGACGGCAACATCCT
EGFP site 69B	TCAGCTCGATGCGGTTCACCagg	gTCAGCTCGATGCGGTTCACC	AGCTGACTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCTCAGCTCGATGCGGTTCACC	TCAGCTCGATGCGGTTCACC
EGFP site 71B	TTCAAGTCCGCCATGCCCGAagg	gTTCAAGTCCGCCATGCCCGA	CTTGAACTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCTTCAAGTCCGCCATGCCCGA	TTCAAGTCCGCCATGCCCGA
EGFP site 72B	TGAAGAAGATGGTGCGCTCCtgg	gTGAAGAAGATGGTGCGCTCC	TCTTCACTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCTGAAGAAGATGGTGCGCTCC	TGAAGAAGATGGTGCGCTCC
EGFP site 73B	TGTACTCCAGCTTGTGCCCCagg	gTGTACTCCAGCTTGTGCCCC	AGTACACTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCTGTACTCCAGCTTGTGCCCC	TGTACTCCAGCTTGTGCCCC
EGFP site 74B	TGCCGTCCTCGATGTTGTGGcgg	gTGCCGTCCTCGATGTTGTGG	ACGGCACTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCTGCCGTCCTCGATGTTGTGG	TGCCGTCCTCGATGTTGTGG
EGFP site 75B	CCGCGCCGAGGTGAAGTTCGagg	gCCGCGCCGAGGTGAAGTTCG	GCGCGGCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCCCGCGCCGAGGTGAAGTTCG	CCGCGCCGAGGTGAAGTTCG

Appendix Table 9. List of target sites and spacers used for the comparison of 21G-sgRNAs, ribosyme flanked-sgRNAs, tRNA flanked-sgRNAs .

References

1 Ginn, S. L., Amaya, A. K., Alexander, I. E., Edelstein, M. & Abedi, M. R. Gene therapy clinical trials worldwide to 2017: An update. *J Gene Med* **20**, e3015, doi:10.1002/jgm.3015 (2018).

2 Cyranoski, D. CRISPR gene-editing tested in a person for the first time. *Nature* **539**, 479, doi:10.1038/nature.2016.20988 (2016).

3 Dunbar, C. E. *et al.* Gene therapy comes of age. *Science* **359**, doi:10.1126/science.aan4672 (2018).

4 Salmaninejad, A. *et al.* Duchenne muscular dystrophy: an updated review of common available therapies. *Int J Neurosci* **128**, 854-864, doi:10.1080/00207454.2018.1430694 (2018).

5 Memi, F., Ntokou, A. & Papangeli, I. CRISPR/Cas9 gene-editing: Research technologies, clinical applications and ethical considerations. *Semin Perinatol* **42**, 487-500, doi:10.1053/j.semperi.2018.09.003 (2018).

6 Hong, A. CRISPR in personalized medicine: Industry perspectives in gene editing. *Semin Perinatol* **42**, 501-507, doi:10.1053/j.semperi.2018.09.008 (2018).

7 Gao, C. The future of CRISPR technologies in agriculture. *Nat Rev Mol Cell Biol* **19**, 275-276, doi:10.1038/nrm.2018.2 (2018).

8 Courtier-Orgogozo, V., Morizot, B. & Boete, C. Agricultural pest control with CRISPR-based gene drive: time for public debate: Should we use gene drive for pest control? *EMBO Rep* **18**, 878-880, doi:10.15252/embr.201744205 (2017).

9 Ricroch, A. Global developments of genome editing in agriculture. *Transgenic Res* **28**, 45-52, doi:10.1007/s11248-019-00133-6 (2019).

10 Mali, P., Esvelt, K. M. & Church, G. M. Cas9 as a versatile tool for engineering biology. *Nat Methods* **10**, 957-963, doi:10.1038/nmeth.2649 (2013).

11 Mali, P. *et al.* CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol* **31**, 833-838, doi:10.1038/nbt.2675 (2013).

12 Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol* **32**, 347-355, doi:10.1038/nbt.2842 (2014).

13 Komor, A. C., Badran, A. H. & Liu, D. R. CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. *Cell* **169**, 559, doi:10.1016/j.cell.2017.04.005 (2017).

14 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, doi:10.1038/nature17946 (2016).

15 Guha, T. K., Wai, A. & Hausner, G. Programmable Genome Editing Tools and their Regulation for Efficient Genome Engineering. *Comput Struct Biotechnol J* **15**, 146-160, doi:10.1016/j.csbj.2016.12.006 (2017).

16 Gordon, J. W., Scangos, G. A., Plotkin, D. J., Barbosa, J. A. & Ruddle, F. H. Genetic transformation of mouse embryos by microinjection of purified DNA. *Proc Natl Acad Sci U S A* **77**, 7380-7384, doi:10.1073/pnas.77.12.7380 (1980).

17 Smithies, O., Gregg, R. G., Boggs, S. S., Koralewski, M. A. & Kucherlapati, R. S. Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. *Nature* **317**, 230-234, doi:10.1038/317230a0 (1985).

18 Lieber, M. R. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu Rev Biochem* **79**, 181-211, doi:10.1146/annurev.biochem.052308.093131 (2010).

19 Zaboikin, M., Zaboikina, T., Freter, C. & Srinivasakumar, N. Non-Homologous End Joining and Homology Directed DNA Repair Frequency of Double-Stranded Breaks Introduced by Genome Editing Reagents. *PLoS One* **12**, e0169931, doi:10.1371/journal.pone.0169931 (2017).

20 San Filippo, J., Sung, P. & Klein, H. Mechanism of eukaryotic homologous recombination. *Annu Rev Biochem* **77**, 229-257, doi:10.1146/annurev.biochem.77.061306.125255 (2008).

21 Rudin, N., Sugarman, E. & Haber, J. E. Genetic and physical analysis of double-strand break repair and recombination in Saccharomyces cerevisiae. *Genetics* **122**, 519-534 (1989).

22 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, doi:10.1128/mcb.14.12.8096 (1994).

23 Smith, H. O. & Wilcox, K. W. A restriction enzyme from Hemophilus influenzae. I. Purification and general properties. 1970. *Biotechnology* **24**, 38-50 (1992).

24 Kelly, T. J., Jr. & Smith, H. O. A restriction enzyme from Hemophilus influenzae. II. *J Mol Biol* **51**, 393-409, doi:10.1016/0022-2836(70)90150-6 (1970).

25 Danna, K. & Nathans, D. Specific cleavage of simian virus 40 DNA by restriction endonuclease of Hemophilus influenzae. *Proc Natl Acad Sci U S A* **68**, 2913-2917, doi:10.1073/pnas.68.12.2913 (1971).

26 Chandrasegaran, S. & Carroll, D. Origins of Programmable Nucleases for Genome Engineering. *J Mol Biol* **428**, 963-989, doi:10.1016/j.jmb.2015.10.014 (2016).

27 Adli, M. The CRISPR tool kit for genome editing and beyond. *Nat Commun* **9**, 1911, doi:10.1038/s41467-018-04252-2 (2018).

28 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, doi:10.1073/pnas.93.3.1156 (1996).

29 Christian, M. *et al.* Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* **186**, 757-761, doi:10.1534/genetics.110.120717 (2010).

30 Carroll, D. Genome Editing: Past, Present, and Future. *Yale J Biol Med* **90**, 653-659 (2017).

31 LaFountaine, J. S., Fathe, K. & Smyth, H. D. Delivery and therapeutic applications of gene editing technologies ZFNs, TALENs, and CRISPR/Cas9. *Int J Pharm* **494**, 180-194, doi:10.1016/j.ijpharm.2015.08.029 (2015).

32 Tebas, P. *et al.* Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. *N Engl J Med* **370**, 901-910, doi:10.1056/NEJMoa1300662 (2014).

33 Reardon, S. Leukaemia success heralds wave of gene-editing therapies. *Nature* **527**, 146-147, doi:10.1038/nature.2015.18737 (2015).

34 Qasim, W. *et al.* Molecular remission of infant B-ALL after infusion of universal TALEN gene-edited CAR T cells. *Sci Transl Med* **9**, doi:10.1126/scitranslmed.aaj2013 (2017).

35 Jinek, M. *et al.* RNA-programmed genome editing in human cells. *Elife* **2**, e00471, doi:10.7554/eLife.00471 (2013).

36 Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823, doi:10.1126/science.1231143 (2013).

37 Cox, D. B. T. *et al.* RNA editing with CRISPR-Cas13. *Science* **358**, 1019-1027, doi:10.1126/science.aaq0180 (2017).

38 Barrangou, R. The roles of CRISPR-Cas systems in adaptive immunity and beyond. *Curr Opin Immunol* **32**, 36-41, doi:10.1016/j.coi.2014.12.008 (2015).

39 Faure, G. *et al.* CRISPR-Cas in mobile genetic elements: counter-defence and beyond. *Nat Rev Microbiol* **17**, 513-525, doi:10.1038/s41579-019-0204-7 (2019).

40 Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol* **13**, 722-736, doi:10.1038/nrmicro3569 (2015).

41 Koonin, E. V., Makarova, K. S. & Zhang, F. Diversity, classification and evolution of CRISPR-Cas systems. *Curr Opin Microbiol* **37**, 67-78, doi:10.1016/j.mib.2017.05.008 (2017).

42 Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Classification and Nomenclature of CRISPR-Cas Systems: Where from Here? *CRISPR J* **1**, 325-336, doi:10.1089/crispr.2018.0033 (2018).

43 Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821, doi:10.1126/science.1225829 (2012).

44 Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823-826, doi:10.1126/science.1232033 (2013).

45 Muller, M. *et al.* Streptococcus thermophilus CRISPR-Cas9 Systems Enable Specific Editing of the Human Genome. *Mol Ther* **24**, 636-644, doi:10.1038/mt.2015.218 (2016).

46 Ran, F. A. *et al.* In vivo genome editing using Staphylococcus aureus Cas9. *Nature* **520**, 186-191, doi:10.1038/nature14299 (2015).

47 Hou, Z. *et al.* Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis. *Proc Natl Acad Sci U S A* **110**, 15644-15649, doi:10.1073/pnas.1313587110 (2013).

48 Zetsche, B. *et al.* Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759-771, doi:10.1016/j.cell.2015.09.038 (2015).

49 Toth, E. *et al.* Cpf1 nucleases demonstrate robust activity to induce DNA modification by exploiting homology directed repair pathways in mammalian cells. *Biol Direct* **11**, 46, doi:10.1186/s13062-016-0147-0 (2016).

50 Tu, M. *et al.* A 'new lease of life': FnCpf1 possesses DNA cleavage activity for genome editing in human cells. *Nucleic Acids Res* **45**, 11295-11304, doi:10.1093/nar/gkx783 (2017).

51 Toth, E. *et al.* Mb- and FnCpf1 nucleases are active in mammalian cells: activities and PAM preferences of four wild-type Cpf1 nucleases and of their altered PAM specificity variants. *Nucleic Acids Res* **46**, 10272-10285, doi:10.1093/nar/gky815 (2018).

52 Burstein, D. *et al.* New CRISPR-Cas systems from uncultivated microbes. *Nature* **542**, 237-241, doi:10.1038/nature21059 (2017).

53 Liu, J. J. *et al.* CasX enzymes comprise a distinct family of RNA-guided genome editors. *Nature* **566**, 218-223, doi:10.1038/s41586-019-0908-x (2019).

54 Cho, S. W., Kim, S., Kim, J. M. & Kim, J. S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol* **31**, 230-232, doi:10.1038/nbt.2507 (2013).

55 Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262-1278, doi:10.1016/j.cell.2014.05.010 (2014).

56 Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S. & Vale, R. D. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* **159**, 635-646, doi:10.1016/j.cell.2014.09.039 (2014).

57 Zuris, J. A. *et al.* Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nat Biotechnol* **33**, 73-80, doi:10.1038/nbt.3081 (2015).

58 Vora, S., Tuttle, M., Cheng, J. & Church, G. Next stop for the CRISPR revolution: RNA-guided epigenetic regulators. *FEBS J* **283**, 3181-3193, doi:10.1111/febs.13768 (2016).

59 Gaudelli, N. M. *et al.* Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. *Nature* **551**, 464-471, doi:10.1038/nature24644 (2017).

60 Wu, W. Y., Lebbink, J. H. G., Kanaar, R., Geijsen, N. & van der Oost, J. Genome editing by natural and engineered CRISPR-associated nucleases. *Nat Chem Biol* **14**, 642-651, doi:10.1038/s41589-018-0080-x (2018).

61 Zafra, M. P. *et al.* Optimized base editors enable efficient editing in cells, organoids and mice. *Nat Biotechnol* **36**, 888-893, doi:10.1038/nbt.4194 (2018).

62 Grunewald, J. *et al.* Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature* **569**, 433-437, doi:10.1038/s41586-019-1161-z (2019).

63 Garneau, J. E. *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67-71, doi:10.1038/nature09523 (2010).

64 Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569-573, doi:10.1038/nature13579 (2014).

65 Jinek, M. *et al.* Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343**, 1247997, doi:10.1126/science.1247997 (2014).

66 Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935-949, doi:10.1016/j.cell.2014.02.001 (2014).

67 Jiang, F., Zhou, K., Ma, L., Gressel, S. & Doudna, J. A. STRUCTURAL BIOLOGY. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science* **348**, 1477-1481, doi:10.1126/science.aab1452 (2015).

68 Anders, C., Bargsten, K. & Jinek, M. Structural Plasticity of PAM Recognition by Engineered Variants of the RNA-Guided Endonuclease Cas9. *Molecular cell* **61**, 895-902, doi:10.1016/j.molcel.2016.02.020 (2016).

69 Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733-740, doi:10.1099/mic.0.023960-0 (2009).

70 Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature biotechnology* **31**, 827-832, doi:10.1038/nbt.2647 (2013).

71 Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature biotechnology* **31**, 822-826, doi:10.1038/nbt.2623 (2013).

72 Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nature biotechnology* **31**, 839-843, doi:10.1038/nbt.2673 (2013).

73 Cradick, T. J., Fine, E. J., Antico, C. J. & Bao, G. CRISPR/Cas9 systems targeting beta-globin and CCR5 genes have substantial off-target activity. *Nucleic acids research* **41**, 9584-9592, doi:10.1093/nar/gkt714 (2013).

74 Cho, S. W. *et al.* Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res* **24**, 132-141, doi:10.1101/gr.162339.113 (2014).

75 Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol* **33**, 187-197, doi:10.1038/nbt.3117 (2015).

76 Frock, R. L. *et al.* Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nature biotechnology* **33**, 179-186, doi:10.1038/nbt.3101 (2015).

77 Wang, X. *et al.* Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. *Nat Biotechnol* **33**, 175-178, doi:10.1038/nbt.3127 (2015).

78 Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M. & Joung, J. K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol* **32**, 279-284, doi:10.1038/nbt.2808 (2014).

79 Tsai, S. Q. & Joung, J. K. Defining and improving the genome-wide specificities of CRISPR-Cas9 nucleases. *Nat Rev Genet* **17**, 300-312, doi:10.1038/nrg.2016.28 (2016).

80 Vouillot, L., Thelie, A. & Pollet, N. Comparison of T7E1 and surveyor mismatch cleavage assays to detect mutations triggered by engineered nucleases. *G3 (Bethesda)* **5**, 407-415, doi:10.1534/g3.114.015834 (2015).

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

82 Smith, C. *et al.* Whole-genome sequencing analysis reveals high specificity of CRISPR/Cas9 and TALEN-based genome editing in human iPSCs. *Cell Stem Cell* **15**, 12-13, doi:10.1016/j.stem.2014.06.011 (2014).

83 Tsai, S. Q. & Joung, J. K. Defining and improving the genome-wide specificities of CRISPR-Cas9 nucleases. *Nat Rev Genet* **17**, 300-312, doi:10.1038/nrg.2016.28 (2016).

84 Kim, D. *et al.* Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat Methods* **12**, 237-243, 231 p following 243, doi:10.1038/nmeth.3284 (2015).

85 Cameron, P. *et al.* Mapping the genomic landscape of CRISPR-Cas9 cleavage. *Nat Methods* **14**, 600-606, doi:10.1038/nmeth.4284 (2017).

86 Tsai, S. Q. *et al.* CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets. *Nat Methods* **14**, 607-614, doi:10.1038/nmeth.4278 (2017).

87 Gabriel, R. *et al.* An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat Biotechnol* **29**, 816-823, doi:10.1038/nbt.1948 (2011).

88 Chiarle, R. *et al.* Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. *Cell* **147**, 107-119, doi:10.1016/j.cell.2011.07.049 (2011).

89 Yan, W. X. *et al.* BLISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks. *Nat Commun* **8**, 15058, doi:10.1038/ncomms15058 (2017).

90 Wienert, B. *et al.* Unbiased detection of CRISPR off-targets in vivo using DISCOVER-Seq. *Science* **364**, 286-289, doi:10.1126/science.aav9023 (2019).

91 Kleinstiver, B. P. *et al.* Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481-485, doi:10.1038/nature14592 (2015).

92 Nishimasu, H. *et al.* Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science* **361**, 1259-1262, doi:10.1126/science.aas9129 (2018).

93 Hu, J. H. *et al.* Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57-63, doi:10.1038/nature26155 (2018).

94 Zetsche, B., Volz, S. E. & Zhang, F. A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nature biotechnology* **33**, 139-142, doi:10.1038/nbt.3149 (2015).

95 Davis, K. M., Pattanayak, V., Thompson, D. B., Zuris, J. A. & Liu, D. R. Small molecule-triggered Cas9 protein with improved genome-editing specificity. *Nature chemical biology* **11**, 316-318, doi:10.1038/nchembio.1793 (2015).

96 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, doi:10.1101/gr.171322.113 (2014).

97 Fu, Y., Reyon, D. & Joung, J. K. Targeted genome editing in human cells using CRISPR/Cas nucleases and truncated guide RNAs. *Methods Enzymol* **546**, 21-45, doi:10.1016/B978-0-12-801185-0.00002-7 (2014).

98 Kocak, D. D. *et al.* Increasing the specificity of CRISPR systems with engineered RNA secondary structures. *Nat Biotechnol*, doi:10.1038/s41587-019-0095-1 (2019).

99 Ran, F. A. *et al.* Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154**, 1380-1389, doi:10.1016/j.cell.2013.08.021 (2013).

100 Tsai, S. Q. *et al.* Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat Biotechnol* **32**, 569-576, doi:10.1038/nbt.2908 (2014).

101 Guilinger, J. P., Thompson, D. B. & Liu, D. R. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat Biotechnol* **32**, 577-582, doi:10.1038/nbt.2909 (2014).

102 Wyvekens, N., Topkar, V. V., Khayter, C., Joung, J. K. & Tsai, S. Q. Dimeric CRISPR RNA-Guided FokI-dCas9 Nucleases Directed by Truncated gRNAs for Highly Specific Genome Editing. *Hum Gene Ther* **26**, 425-431, doi:10.1089/hum.2015.084 (2015).

103 Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84-88, doi:10.1126/science.aad5227 (2016).

104 Kleinstiver, B. P. *et al.* High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490-495, doi:10.1038/nature16526 (2016).

105 Crosetto, N. *et al.* Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. *Nature methods* **10**, 361-365, doi:10.1038/nmeth.2408 (2013).

106 Chen, J. S. *et al.* Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* **550**, 407-410, doi:10.1038/nature24268 (2017).

107 Casini, A. *et al.* A highly specific SpCas9 variant is identified by in vivo screening in yeast. *Nat Biotechnol* **36**, 265-271, doi:10.1038/nbt.4066 (2018).

108 Goomer, R. S. & Kunkel, G. R. The transcriptional start site for a human U6 small nuclear RNA gene is dictated by a compound promoter element consisting of the PSE and the TATA box. *Nucleic Acids Res* **20**, 4903-4912 (1992).

109 Paule, M. R. & White, R. J. Survey and summary: transcription by RNA polymerases I and III. *Nucleic Acids Res* **28**, 1283-1298, doi:10.1093/nar/28.6.1283 (2000).

110 Miyagishi, M. & Taira, K. U6 promoter-driven siRNAs with four uridine 3' overhangs efficiently suppress targeted gene expression in mammalian cells. *Nat Biotechnol* **20**, 497-500, doi:10.1038/nbt0502-497 (2002).

111 Milligan, J. F., Groebe, D. R., Witherell, G. W. & Uhlenbeck, O. C. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res* **15**, 8783-8798 (1987).

112 Beckert, B. & Masquida, B. Synthesis of RNA by in vitro transcription. *Methods Mol Biol* **703**, 29-41, doi:10.1007/978-1-59745-248-9_3 (2011).

113 Moreno-Mateos, M. A. *et al.* CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nat Methods* **12**, 982-988, doi:10.1038/nmeth.3543 (2015).

114 Kunkel, G. R., Maser, R. L., Calvet, J. P. & Pederson, T. U6 small nuclear RNA is transcribed by RNA polymerase III. *Proc Natl Acad Sci U S A* **83**, 8575-8579, doi:10.1073/pnas.83.22.8575 (1986).

115 Nowak, C. M., Lawson, S., Zerez, M. & Bleris, L. Guide RNA engineering for versatile Cas9 functionality. *Nucleic Acids Res* **44**, 9555-9564, doi:10.1093/nar/gkw908 (2016).

116 Ma, H. *et al.* Pol III Promoters to Express Small RNAs: Delineation of Transcription Initiation. *Mol Ther Nucleic Acids* **3**, e161, doi:10.1038/mtna.2014.12 (2014).

117 Gao, Z., Harwig, A., Berkhout, B. & Herrera-Carrillo, E. Mutation of nucleotides around the +1 position of type 3 polymerase III promoters: The effect on transcriptional activity and start site usage. *Transcription* **8**, 275-287, doi:10.1080/21541264.2017.1322170 (2017).

118 DeWitt, M. A. *et al.* Selection-free genome editing of the sickle mutation in human adult hematopoietic stem/progenitor cells. *Sci Transl Med* **8**, 360ra134, doi:10.1126/scitranslmed.aaf9336 (2016).

119 Rees, H. A. *et al.* Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. *Nat Commun* **8**, 15790, doi:10.1038/ncomms15790 (2017).

120 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, doi:10.1038/s41591-018-0137-0 (2018).

121 Lee, J. K. *et al.* Directed evolution of CRISPR-Cas9 to increase its specificity. *Nat Commun* **9**, 3048, doi:10.1038/s41467-018-05477-x (2018).

122 Engler, C., Kandzia, R. & Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PloS one* **3**, e3647 (2008).

123 Kostylev, M., Otwell, A. E., Richardson, R. E. & Suzuki, Y. Cloning Should Be Simple: Escherichia coli DH5 alpha-Mediated Assembly of Multiple DNA Fragments with Short End Homologies. *Plos One* **10**, doi:ARTN e013746610.1371/journal.pone.0137466 (2015).

124 Tóth, E. *et al.* Restriction enzyme body doubles and PCR cloning: on the general use of type IIs restriction enzymes for cloning. *PloS one* **9** (2014).

125 Iwamoto, M., Bjorklund, T., Lundberg, C., Kirik, D. & Wandless, T. J. A general chemical method to regulate protein stability in the mammalian central

nervous system. *Chem Biol* **17**, 981-988, doi:10.1016/j.chembiol.2010.07.009 (2010).

126 Chavez, A. *et al.* Highly efficient Cas9-mediated transcriptional programming. *Nature methods* **12**, 326-328, doi:10.1038/nmeth.3312 (2015).

127 Gilbert, L. A. *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442-451, doi:10.1016/j.cell.2013.06.044 (2013).

128 Perez-Pinera, P. *et al.* RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nature methods* **10**, 973-976, doi:10.1038/nmeth.2600 (2013).

129 Maeder, M. L. *et al.* CRISPR RNA-guided activation of endogenous human genes. *Nature methods* **10**, 977-979, doi:10.1038/nmeth.2598 (2013).

130 Chavez, A. *et al.* Comparison of Cas9 activators in multiple species. *Nature methods* **13**, 563-567, doi:10.1038/nmeth.3871 (2016).

131 Gilbert, L. A. *et al.* Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* **159**, 647-661, doi:10.1016/j.cell.2014.09.029 (2014).

132 Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173-1183, doi:10.1016/j.cell.2013.02.022 (2013).

133 Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583-588, doi:10.1038/nature14136 (2015).

134 Peabody, D. S. The RNA binding site of bacteriophage MS2 coat protein. *EMBO J* **12**, 595-600 (1993).

135 Vriend, L. E., Jasin, M. & Krawczyk, P. M. Assaying break and nick-induced homologous recombination in mammalian cells using the DR-GFP reporter and Cas9 nucleases. *Methods Enzymol* **546**, 175-191, doi:10.1016/B978-0-12-801185-0.00009-X (2014).

136 Tsai, S. Q., Topkar, V. V., Joung, J. K. & Aryee, M. J. Open-source guideseq software for analysis of GUIDE-seq data. *Nat Biotechnol* **34**, 483, doi:10.1038/nbt.3534 (2016).

137 Talas, A. *et al.* A convenient method to pre-screen candidate guide RNAs for CRISPR/Cas9 gene editing by NHEJ-mediated integration of a 'self-cleaving' GFP-expression plasmid. *DNA Res* **24**, 609-621, doi:10.1093/dnares/dsx029 (2017).

138 Kulcsar, P. I. *et al.* Crossing enhanced and high fidelity SpCas9 nucleases to optimize specificity and cleavage. *Genome Biol* **18**, 190, doi:10.1186/s13059-017-1318-8 (2017).

139 Ma, H. *et al.* A CRISPR-Based Screen Identifies Genes Essential for West-Nile-Virus-Induced Cell Death. *Cell Rep* **12**, 673-683, doi:10.1016/j.celrep.2015.06.049 (2015).

140 Koike-Yusa, H., Li, Y., Tan, E. P., Velasco-Herrera Mdel, C. & Yusa, K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat Biotechnol* **32**, 267-273, doi:10.1038/nbt.2800 (2014).

141 Tzelepis, K. *et al.* A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets in Acute Myeloid Leukemia. *Cell Rep* **17**, 1193-1205, doi:10.1016/j.celrep.2016.09.079 (2016).

142 Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* **34**, 184-191, doi:10.1038/nbt.3437 (2016).

143 Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat Methods* **11**, 783-784, doi:10.1038/nmeth.3047 (2014).

144 Morgens, D. W. *et al.* Genome-scale measurement of off-target activity using Cas9 toxicity in high-throughput screens. *Nat Commun* **8**, 15178, doi:10.1038/ncomms15178 (2017).

145 Gao, Y. & Zhao, Y. Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing. *J Integr Plant Biol* **56**, 343-349, doi:10.1111/jipb.12152 (2014).

146 Kim, S., Bae, T., Hwang, J. & Kim, J. S. Rescue of high-specificity Cas9 variants using sgRNAs with matched 5' nucleotides. *Genome Biol* **18**, 218, doi:10.1186/s13059-017-1355-3 (2017).

147 Lee, R. T. H., Ng, A. S. M. & Ingham, P. W. Ribozyme Mediated gRNA Generation for In Vitro and In Vivo CRISPR/Cas9 Mutagenesis. *Plos One* **11**, doi:ARTN e016602010.1371/journal.pone.0166020 (2016).

148 Nissim, L., Perli, S. D., Fridkin, A., Perez-Pinera, P. & Lu, T. K. Multiplexed and programmable regulation of gene networks with an integrated RNA and CRISPR/Cas toolkit in human cells. *Mol Cell* **54**, 698-710, doi:10.1016/j.molcel.2014.04.022 (2014).

149 Xie, K., Minkenberg, B. & Yang, Y. Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proc Natl Acad Sci U S A* **112**, 3570-3575, doi:10.1073/pnas.1420294112 (2015).

150 Dong, F., Xie, K., Chen, Y., Yang, Y. & Mao, Y. Polycistronic tRNA and CRISPR guide-RNA enables highly efficient multiplexed genome engineering in human cells. *Biochem Biophys Res Commun* **482**, 889-895, doi:10.1016/j.bbrc.2016.11.129 (2017).

151 Zhang, D. *et al.* Perfectly matched 20-nucleotide guide RNA sequences enable robust genome editing using high-fidelity SpCas9 nucleases. *Genome Biol* **18**, 191, doi:10.1186/s13059-017-1325-9 (2017).

152 Spitzer, M., Wildenhain, J., Rappsilber, J. & Tyers, M. BoxPlotR: a web tool for generation of box plots. *Nat Methods* **11**, 121-122, doi:10.1038/nmeth.2811 (2014).

153 Kiani, S. *et al.* Cas9 gRNA engineering for genome editing, activation and repression. *Nature methods* **12**, 1051-1054, doi:10.1038/nmeth.3580 (2015).

154 Dahlman, J. E. *et al.* Orthogonal gene knockout and activation with a catalytically active Cas9 nuclease. *Nature biotechnology* **33**, 1159-1161, doi:10.1038/nbt.3390 (2015).

155 Jiang, F. G. *et al.* Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* **351**, 867-871, doi:10.1126/science.aad8282 (2016).

156 Jiang, F. *et al.* Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* **351**, 867-871, doi:10.1126/science.aad8282 (2016).

157 Lee, R. T., Ng, A. S. & Ingham, P. W. Ribozyme Mediated gRNA Generation for In Vitro and In Vivo CRISPR/Cas9 Mutagenesis. *PLoS One* **11**, e0166020, doi:10.1371/journal.pone.0166020 (2016).

158 Sternberg, S. H., LaFrance, B., Kaplan, M. & Doudna, J. A. Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature* **527**, 110-113, doi:10.1038/nature15544 (2015).

159 Dagdas, Y. S., Chen, J. S., Sternberg, S. H., Doudna, J. A. & Yildiz, A. A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. *Sci Adv* **3**, eaao0027, doi:10.1126/sciadv.aao0027 (2017).

160 Singh, D., Sternberg, S. H., Fei, J., Doudna, J. A. & Ha, T. Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9. *Nat Commun* **7**, 12778, doi:10.1038/ncomms12778 (2016).

161 Ryan, D. E. *et al.* Improving CRISPR-Cas specificity with chemical modifications in single-guide RNAs. *Nucleic Acids Res* **46**, 792-803, doi:10.1093/nar/gkx1199 (2018).

162 Yin, H. *et al.* Partial DNA-guided Cas9 enables genome editing with reduced off-target activity. *Nat Chem Biol* **14**, 311-316, doi:10.1038/nchembio.2559 (2018).

163 Ricci, C. G. *et al.* Deciphering Off-Target Effects in CRISPR-Cas9 through Accelerated Molecular Dynamics. *ACS Cent Sci* **5**, 651-662, doi:10.1021/acscentsci.9b00020 (2019).

164 Nishimasu, H. *et al.* Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science*, doi:10.1126/science.aas9129 (2018).

I

Társszerzői lemondó nyilatkozat

Alulírott Dr. Welker Ervin (felelős szerző) kijelentem, hogy Kulcsár Péter István (jelölt) PhD értekezésében bemutatott tudományos eredmények a jelölt saját eredményei (amennyiben nincsen külön jelölve, hogy a kísérletet más végezte el), valamint a pályázónak meghatározó szerepe volt a kézirat elkészítésében, ezért ezeket az eredményeket más nem használta fel és a jövőben sem kívánja felhasználni PhD fokozat szerzésre.

…….

Dr. Welker Ervin

Budapest, 2019.06.12

Kulcsár P. I., Tálas A., Huszár K., Ligeti Z., Tóth E., Weinhardt N., Fodor E., Welker E., Crossing enhanced and high fidelity SpCas9 nucleases to optimize specificity and cleavage, Genome Biology (2017).

II

Társszerzői lemondó nyilatkozat

Alulírott Tálas András (első szerző) kijelentem, hogy a cikkben a Figure S6 ábrán bemutatott eredmények Kulcsár Péter István, társszerző munkája. A cikk ezen eredményeit sem én, sem más nem használta fel és a jövőben sem kívánja felhasználni fokozatszerzésre.

…….

Tálas András

Budapest, 2019.06.12

Tálas A., **Kulcsár P. I.**, Weinhardt N., Borsy A., Tóth E., Szebényi K., Krausz S. L., Huszár K., Vida I., Sturm Á., Gordos B., Hoffmann O. I., Bencsura P., Nyeste A., Ligeti Z., Fodor E., Welker E., A convenient method to pre-screen candidate guide RNAs for CRISPR/Cas9 gene editing by NHEJ-mediated integration of a ‘self-cleaving’ GFP-expression plasmid, DNA Research (2017)