



UNIVERSITY OF TRENTO - Italy

PhD Program in Biomolecular Sciences
Centre for Integrative Biology
27th Cycle

**“Better safe than sorry: new CRISPR/Cas9 tools
for improved genome engineering”**

Tutor

Anna CERESETO

CIBIO, University of Trento

Ph.D. Thesis of

Antonio CASINI

CIBIO, University of Trento

Academic Year 2015-2016

Declaration of Authorship

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

Trento, 3rd March 2017

Antonio Casini

To Veronica and my family, which are one

Aim of the thesis	1
Abstract.....	2
Introduction.....	3
<i>The modification of genomes</i>	3
<i>Gene therapy, gene targeting and genome engineering</i>	5
<i>Targeted nucleases</i>	11
<i>The CRISPR/Cas system</i>	20
<i>CRISPR type II systems: the Cas9 endonuclease</i>	24
<i>Biotechnological applications of CRISPR nucleases</i>	27
<i>Limitations in the CRISPR/Cas9 technology</i>	32
<i>Other recently discovered editing systems</i>	34
<i>Safety issues and off-target activity</i>	35
<i>Methods to reduce Cas9 off-target activity</i>	42
Identification of high-fidelity Cas9 variants using a yeast-based screening.....	49
<i>Rationale of the experiments</i>	49
<i>Methods</i>	51
<i>Results</i>	62
Design of a reporter yeast strain for the detection of Cas9 activity.....	62
Yeast-based screening for high-specificity SpCas9 variants.....	66
Optimization of high-fidelity SpCas9 variants in mammalian cells.....	69
Characterization of sgRNA requirements of evoCas9.....	72
evoCas9 activity towards endogenous loci.....	77
Evaluation of evoCas9 off-target activity.....	78
Genome-wide specificity of evoCas9.....	81
Specificity of an evo-dCas9-based transcriptional activator.....	82
Long-term specificity of evoCas9.....	84
Hit-and-go Cas9 delivered through a lentiviral based self-limiting circuit.....	87
<i>Rationale of the experiments</i>	87
<i>Methods</i>	88
<i>Results</i>	93
Effects of long-term Cas9 expression on editing activity.....	93
Design of a Self-Limiting Cas9 circuitry for Enhanced Safety (SLiCES).....	94
Effect of SLiCES on gene editing by homologous recombination.....	99
Adaptation of SLiCES to other RNA-guided nucleases.....	101
SLiCES specificity towards endogenous genomic loci.....	102
Adaptation of SLiCES to lentiviral delivery.....	103
Characterization of lentiSLiCES specificity.....	105
Discussion.....	108
Contribution to the experiments.....	113
Acknowledgments.....	114
Bibliography.....	116
Appendix.....	133

Aim of the thesis

A major need in the genome editing community is the generation of an error-free toolbox to allow its *in vivo* application for therapeutic purposes. The aim of this thesis is the development and the validation of innovative technologies to reduce SpCas9 off-target activity. We addressed critical aspects related to Cas9 specificity as well as important issues connected to the accumulation of unwanted cleavages after long term expression of the nuclease into cells.

The first goal of the present study is the identification of high-fidelity SpCas9 variants with significantly reduced off-target activity. To this aim we set-up a yeast-based *in vivo* screening platform to allow the unbiased isolation amino acid substitutions increasing SpCas9 targeting specificity. An in depth characterization of the identified variants in terms of specificity, as well as on-target activity, was performed. Further analyses will be required to complete the structural and biochemical characterization of the selected hits.

To address the long-term adverse effects related to Cas9 permanence into cells, we decided to engineer a self-limiting synthetic circuit deliverable through an all-in-one lentiviral vector to switch-off Cas9 expression in transduced cells. This system is designed to generate a peak-like pulse of nuclease activity sufficient to modify the locus of interest, while limiting time- and dose-dependent Cas9 detrimental effects.

Further optimizations of the platforms here described could be envisioned, such as the isolation of alternative high-specificity variants, the application of the *in vivo* screening assay to other RNA guided nucleases or additional implementations of the self-limiting circuit involving other vector systems. The work hereby presented aims at validating these new technologies as valuable tools to prevent Cas9 unwanted effects and anticipates their employment in everyday research protocols and in *in vivo* experimental disease models, as well as their application as prospective therapeutic agents.

Abstract

CRISPR nucleases are efficient tools to edit cellular genomes in a variety of organisms. However, the *in vivo* application of this technology is still severely limited by unwanted genomic cleavages, that are further increased by long-term expression of the nuclease and can lead to unpredictable results.

To address this limitation, we developed a yeast-based assay which allows to simultaneously evaluate the on- and off-target activity towards two engineered genomic targets in order to select optimized *Streptococcus pyogenes* Cas9 (SpCas9) variants. The screening of SpCas9 variants obtained by random mutagenesis of the Rec1-II domain allowed the identification of hits with increased on/off ratios. Through the combination of the identified mutations within a single variant we isolated the best performing nuclease, that we named evoCas9 (evolved Cas9). Side by side analyses with recently reported rationally designed variants demonstrated a significant improvement in fidelity of our evoCas9.

In addition, to control Cas9 persistence into cells over time, we developed a Self-Limiting Cas9 circuitry for Enhanced Safety (SLICES) which consists of an expression unit for SpCas9, a self-targeting sgRNA and a second sgRNA targeting a chosen genomic locus. This self-limiting circuit, by controlling Cas9 levels, results in increased genome editing specificity. For its *in vivo* utilization, we integrated SLICES into a lentiviral delivery system (lentiSLICES) *via* circuit inhibition to achieve viral particle production. Following its delivery into target cells, the lentiSLICES circuit is switched on to edit the intended genomic locus while simultaneously promoting its own neutralization through SpCas9 inactivation.

The two strategies here developed represent complementary approaches to address a major issue in the genome editing field. On one hand, by preserving target cells from residual nuclease activity, our *hit and go* SLICES system increases the safety margins for genome engineering. On the other, if compared to published structure-guided protein engineering approaches, our *in vivo* screening increases the likelihood to identify the best combination of amino acid substitutions for the generation of novel, error-free SpCas9 and could represent a valid strategy to enhance the specificity of other RNA-guided nucleases.

Introduction

The modification of genomes

The alternation of the genome of useful plants and animals, in the form of domestication and selective breeding, is an activity that humans have been performing since early prehistory¹, well before the discovery that DNA is the carrier of genetic information and the development of modern molecular biology techniques (**Fig. 1**). Charles Darwin devoted the entire first chapter of his “On the Origin of Species” to this topic and used the observations made on artificial selection to give shape to his evolutionary theory. Thus, the possibility to modify the genetic characteristics of farm animals and crops, with the consequent alteration of their phenotype, was a need that accompanied humankind from its dawn and shaped dramatically its history.



Figure 1. Domestication of wheat. From left to right: wild wheat and the artificially selected variants einkorn wheat, emmer wheat and common bread wheat. The increased productivity of each ear is immediately apparent. Retrieved from Slideshare.net.

It was not until mid-1800 onwards that the idea of inheritable characters was suggested by Gregor Mendel, giving birth to classical genetics, and only almost a century later DNA was identified as the repository of genetic information². In the following years, the development of the bases of modern molecular biology, that allowed the *in vitro* manipulation of DNA molecules, led to the first experiments with the specific aim of altering the genetic makeup of a lab-grown bacteria³. Only few years later, Rudolf Jaenisch developed the first transgenic mouse⁴ pushing even further the boundaries of artificial genomic modification. The profound change in the ways in which researchers could act on and interact with biological systems was deeply perceived by

the scientific community that felt the need to call for an international meeting with the goal of developing a common regulatory framework on genetic engineering, the Council of Asilomar (1975). The further development of the knowledge on DNA, the possibility to read and decipher the information it contains by ever-evolving sequencing methods and the advent of technological advances that significantly reduced the costs of biotechnological applications led to a massive expansion of the use of genetic engineering, even outside research labs. The introduction of genetic engineering into the industry allowed the production of several genetically modified crops (famous examples are the Flavr Savr tomato and several varieties of rice, soy, cotton, maize, canola) as well as numerous recombinant proteins with a relevant therapeutic impact, such as insulin or the growth hormone, which were previously very difficult or impossible to recover from natural sources.

Recently, the creation of a mycobacterium cell with a completely *in vitro* synthesized genome⁵ and then of an upgraded version of the same cell with a minimal genome containing only a set of essential genes⁶ is an indication of how far the modification of genomes could be pushed: building a new genome from scratch, even though the possibility to start from an engineered blueprint is not a reality yet. While the ability to assemble entire genomes is limited to relatively simple circular prokaryotic DNA molecules, the recent discovery of programmable endonucleases allowed the extensive modification of more complex plant and mammalian genomes and represents the state-of-the-art in terms of genomic engineering in complex organisms. In addition, the ease of use of programmable endonucleases and their low operative costs transformed genome editing into a mass-phenomenon. This, together with the introduction on the market of genetically modified organisms (GMOs), generated a heated debate in the public opinion, that in general looks with suspicion at such artificial manipulations, although sometimes the final result is very similar to what could have been obtained by traditional breeding techniques, such as in the case of crops able to grow in harsh environments⁷.

The different perception that nowadays the non-specialized public has of genetic engineering underlines the fundamental differences between these new technologies and the more traditional ones. The most evident is the speed at which the desired phenotype can be obtained, as targeted genomic modifications are characterized by

an almost immediate effect if compared to the generations of crossings needed to isolate a pure line with the desired characters. Secondly, it is now clear that to modify the phenotype we need to alter the genome of the individual, while in the past only the final effects on the phenotype were perceived. Third, researchers are now able to directly alter the genomic DNA of many organisms rather than indirectly putting a selective pressure on certain characters and leave the rest of the work to evolution. Connected to the last aspect, the idea of swapping pieces of DNA between different species and even between different kingdoms is an unprecedented possibility and is considered by part of the public opinion as irresponsible and unnatural.

Altogether, although the modification of genomes is an activity that humans have been carrying out since millennia, the discoveries made in the last century opened a completely new scenario with remarkable perspectives and almost endless possibilities, very often connected with important ethical issues.

Gene therapy, gene targeting and genome engineering

The discovery that the main characteristics of all biological systems are determined by their genetic makeup, together with the possibility to modify this genetic material, had an immediate impact on biomedical research, as well as on the therapeutic approach to several diseases. The concept that exogenous “good” DNA could be introduced in diseased cells to restore their normal functions (“genes as drugs”) gave birth to gene therapy, which aim is to cure the disease at its genetic roots, without the need of further medical interventions⁸. The main challenges in gene therapy are to deliver correctly the exogenous DNA to target cells, overcoming cellular and tissue barriers, and to obtain the desired expression pattern of the therapeutic molecule in an appropriate number of cells that must persist in the patient, without being cleared by the immune system, in order to eventually transmit the modification to daughter cells⁹. In the past decades several methods have been developed to efficiently deliver therapeutic DNA to target cells, both *in vivo* and *ex vivo*. On one hand, researcher have devised physical and chemical approaches to introduce transgenes into cells¹⁰. These methods usually suffer from low efficiency and are applicable only in certain contexts. Alternatively, the use of viral vectors has found widespread use thanks to their innate

efficiency in penetrating different cell types, even though safety concerns are more pressing since this kind of delivery tools are often based on parental pathogenic viruses¹¹. Several type of viruses have been modified and tested as gene therapy vectors and some of them have been used in clinical trials to treat a wide range of inherited diseases, cancers and chronic infections, albeit with mixed efficacy^{9,12,13}. For example lentiviruses, γ -retroviruses and adeno-associated viruses (AAV) have all been tested in patients⁹. A few have received the approval by regulatory bodies for commercialization (such as Glybera^{14,15} and Strimvelis^{16,17}). Many gene therapy applications require stable expression of the transgene and for this reason the integration of the transferred DNA is generally required. In this process, known as gene insertion and efficiently mediated by retroviral vectors, the exogenous transgenic nucleic acid does not substitute the non-functional endogenous copy but is simply added to the patient's genome in order to provide the missing molecular function or counteract a malfunctioning gene product. However, the present knowledge of vector biology does not allow a precise engineering of their integration preferences and for this reason the semi-random pattern of insertion into the patient's genome can potentially cause insertional mutagenesis, leading in the worst case to the selection of malignant clones in the population of transduced cells^{11,18}. In the past, this has produced dramatic side effects in a clinical trial for the treatment of X-SCID¹⁹, where vector-induced activation of the LMO-2 proto-oncogene caused leukaemia in some of the patients^{20,21}. Technical improvements in vector design allowed the generation of safer next generation gene therapy tools that however still suffered from limitations such as the impossibility to deliver and integrate big transgenes or the inability to cure conditions for which the diseased gene product exerts a dominant effect which cannot be counteracted by gene insertion strategies. In both cases a targeted correction of the mutated endogenous locus would represent a valid therapeutic approach. In conclusion, even though viral vectors are first-choice tools for gene therapy protocols, they still lack the delivery precision that is needed by advanced genomic surgery approaches.

The necessity to alter specific genomic regions, either for research purposes or for therapeutic protocols, led to the development of gene targeting that was pioneered by Mario Capecchi, Martin Evans and Oliver Smithies who were awarded the 2007 Nobel

Prize for Physiology and Medicine "for their discoveries of principles for introducing specific gene modifications in mice by the use of embryonic stem cells". Their merit was to adapt to mammalian cells observations on homologous gene targeting made previously in yeast^{22,23}, generating a technology that today still remains the gold standard to produce transgenic mice²⁴. Gene targeting relies on the possibility to select specific events of homologous recombination between a heterologous DNA molecule containing the desired modification and the genomic sequence of interest (**Fig. 2**)²⁵⁻²⁸. This approach was revolutionary if compared to previous technologies where random mutations were introduced into the genome and the desired modification was then isolated *a posteriori* from a specific phenotype. The latter methodology relies on one hand on the previous knowledge of the phenotypic change to be expected from the desired alteration, while on the other requires extensive experimental work for mutant isolation and phenotype validation, without the assurance to recover the desired genomic modification.

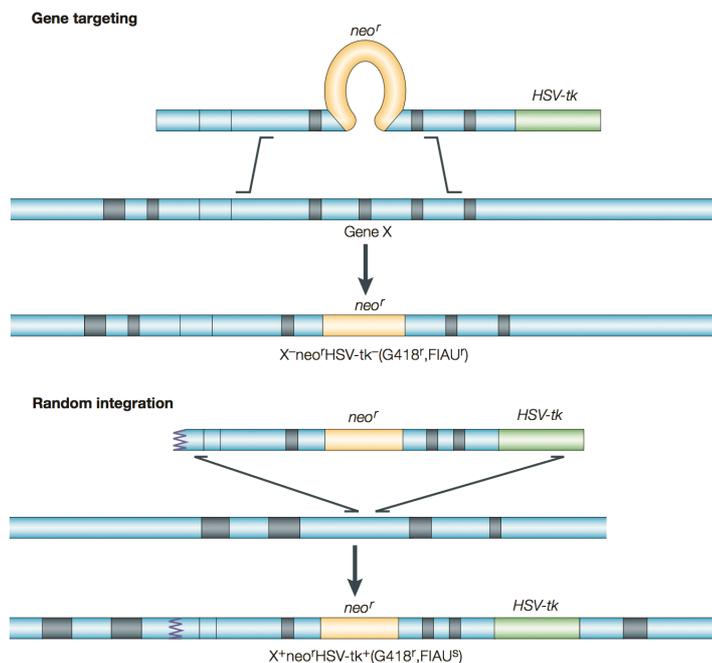


Figure 2. Gene targeting by positive-negative selection. A targeting vector containing appropriate homology arms, the *neo^r* and the HSV thymidine kinase selective markers is introduced into cells. After successful recombination with the Gene X locus, cells acquire G418 resistance while losing the HSV-tk gene. Random integration events that lead to the incorporation of the whole targeting vector into the genome can be counter selected exploiting the HSV-tk. FIAU: fialuridine. Adapted from Capecchi, Nat. Rev. Genet., 2005.

Fundamental to the advancement of the genome modification field was a better understanding of how the cell copes with the presence of exogenous DNA molecules that share some similarities with endogenous sequences. The discovery that cells have an intrinsic ability to target exogenous DNA to chromosomal sites suggested that the cellular machinery responsible for such processes could be harnessed to obtain

targeted modifications of the genome^{22,27}. Moreover, several studies analysed the parameters that influence the recombination reaction, such as the extent of the homology between the exogenous DNA and the target sequence^{29,30}, the type of modifications that are more likely to be introduced into the genome³¹ or the influence of the cell cycle on the efficiency of the process^{32,33}, thus allowing the design of strategies to optimize gene targeting experiments. Even though cells are capable of incorporating exogenous DNA by homologous recombination, the parallel process involving incorporation by non-homologous end joining (NHEJ)³³, which results in a random integration pattern into chromosomes, is actually preferred^{31,34}. For this reason, it is necessary to select cells where this relative rare events took place by using a drug-resistance gene that is inserted in the locus of interest together with the desired DNA sequence. Generally, a second selectable marker is positioned outside the homology regions in order to negatively select cells in which the integration of the targeting construct is mediated by non-homologous end joining, considering that those portions of the construct will be lost due to crossing over in case homologous recombination occurs (**Fig. 2**)^{25,26}. Several variants of the gene targeting technique have been developed over time with the main aim of optimizing the selection procedure to enrich for correctly edited cells; some examples are the "in-out" approach³⁵ or the tag-and-exchange strategy³⁶. The introduction of the Cre/loxP recombinase systems, on the other hand, allowed an easier removal of the selectable markers from the edited locus when needed and, more importantly, granted the possibility of generating conditional knockouts by controlling the expression of the Cre recombinase using inducible or tissue specific promoters or by transfecting or transducing the recombinase locally³⁷.

Despite the progresses made in designing new protocols, gene targeting still suffered from a fundamental limitation that is deeply connected with the principle on which the technique is based: the low frequency at which homologous recombination events take place in the cell. It has been calculated that in higher eukaryotic cells such events have a frequency from 1 in 10^6 to 1 in 10^7 , which renders gene targeting a non-viable option for *in vivo* applications, where selection of edited cells is seldom possible³⁸. The discovery that double strand breaks (DSB) in proximity to the target locus are able to increase the frequency of homologous recombination by several orders of

magnitude³⁹⁻⁴¹, allowed the generation of edited cells with much less effort, if compared to classical gene targeting experiments. This could be considered a milestone towards the possibility to edit the cellular genome *in vivo*, paving the way for innovative therapeutic approaches. As demonstrated by these early studies, where the rare cutting yeast mitochondrial endonuclease I-SceI has been used to generate the DSB at the desired location³⁹⁻⁴¹, the ability to induce DSB at will in a specific genomic location becomes of primary importance. Indeed, the use of endonucleases together with targeting constructs containing specific homology arms allowed the transfer of single-base-pair changes with high efficiency to a mutated chromosomal sequence (**Fig. 3**)⁴². This has clear implications for the correction of disease-causing mutations, as demonstrated by many studies⁴³. In addition, recent works have analysed the characteristic of the donor template, which has been traditionally identified as plasmid DNA with extensive homology with the target locus, suggesting that much shorter single-stranded oligonucleotides⁴⁴⁻⁴⁶ or carefully designed donor molecules⁴⁷ could be more efficient in mediating homologous recombination. In addition, for *in vivo* applications where the delivery of donor molecules could be more difficult and for cells that are simply harder to transfect, the use of integration-defective lentiviral vectors (IDLVs) or AAV vectors as donor substrates has been proposed⁴⁸⁻⁵³.

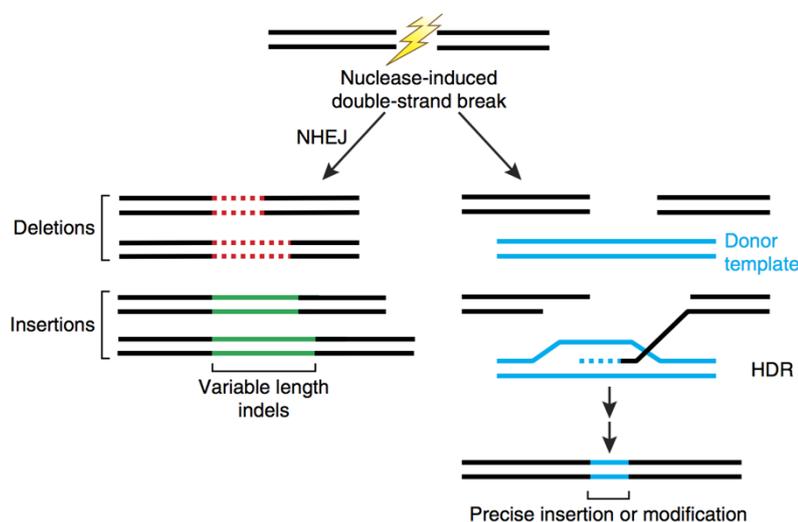


Figure 3. DSB-induced genome modification. Nuclease-induced DSBs can be repaired by the cell either by NHEJ, introducing variable length indels into the target locus, or by homology directed repair, using an exogenous donor template, generating specific modifications to the locus of interest. Adapted from Sanders&Joung, Nat. Biotechnol., 2014.

The recent discovery of programmable nucleases, which will be the topic of the next chapters, for the first time enabled the introduction of DSBs into specific genomic sites, boosting the efficiency of the addition of exogenous sequences at desired loci into the

human genome, using *ad hoc* designed donor molecules⁵⁴. This allowed the insertion of therapeutic sequences into genomic “safe harbours”, in order to avoid the collateral effects observed in the past with randomly integrating vectors, such as insertional mutagenesis^{18,20}, dysregulation of endogenous gene expression and positional effects on transgene expression⁵⁵. Unfortunately, the identification of genomic safe harbours is a complex procedure that requires a profound knowledge of the target locus, which is seldom available. Alternatively, the replacement of genes containing disease-causing mutations by the targeted insertion of a wild-type counterpart can be promoted by exploiting the same basic principles: the clear advantage here is the possibility to maintain the endogenous regulatory elements and the exact location into the genome⁵⁶, even though the insertion of large coding sequences could be difficult due to a decrease in homology-directed repair efficiency.

As mentioned previously, mammalian cells generally tend to repair DSB using non homologous end joining (NHEJ), a process where two DNA ends are directly joined without the need of a guiding template⁵⁷. This repair pathway is intrinsically error-prone as the DNA ends that take part in the joining reaction are often modified by the removal or addition of a variable number of bases before the ligation. An alternative pathway, called microhomology-mediated end joining (MMEJ), uses short homology regions flanking the DSB to favour the annealing of the two DNA ends before repair, generally producing the deletion of a variable stretch of bases⁵⁸. Similarly to targeted integration approaches, the use of programmable nucleases allowed the exploitation of these molecular events to produce gene knockouts with high efficiency, as the indels that would likely result as a byproduct of the repair reaction will lead to frameshift mutations and to premature termination of protein translation if existing out-of-frame stop codon are brought in frame (**Fig. 3**)^{58,59}. It is also possible to couple two DSB to delete the genomic sequence included between the two target sites, removing regions up to some megabases in size⁶⁰⁻⁶³. This strategy is particularly useful when entire intronic or exonic portions of genes have to be excised^{64,65} or when it is necessary to target regulatory elements. The generation of knockouts is extremely useful in research applications, as it allows the dissection of complex biological functions, and in many instances it is superior to RNA-based knockdown approaches, where the downregulation of gene expression is never complete. The knockout technology has

also interesting therapeutic applications such as the removal of gene products with toxic gain of function dominant mutations, like the ones linked with trinucleotide repeats expansion⁶⁶. Notably, one of the first clinical trials using gene editing techniques has the goal of assessing if the knockdown the CCR5 gene in immune cells makes them resistant to HIV-1 infection⁶⁷⁻⁶⁹. In fact, CCR5 is one of HIV-1 co-receptors and is not essential for the normal function of T cells, as demonstrated by the occurrence of individuals with an inactivating deletion into the gene that are naturally resistant to HIV-1 infection⁷⁰.

Altogether, the possibility to modify the genome with relative ease and the recent discovery of several different classes of programmable nucleases that allow the targeting of virtually any genomic locus, gave birth to a whole new field of genome biology: genome engineering.

Targeted nucleases

As outlined in the previous chapter, fundamental for the development of genome editing approaches is the availability of targeted endonucleases to produce DSB at specific locations in the genome in order to stimulate the activation of the endogenous repair machinery either for NHEJ-mediated knockouts or homologous recombination dependent gene editing. To date, four main classes of targeted nucleases have been widely used: meganucleases, zinc finger nucleases (ZFNs), Transcription Activator-Like Effector (TALE) nucleases (TALENs) and the more recently discovered Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/ CRISPR associated (Cas) nucleases, also called RNA guided nucleases (RGNs).

Meganucleases. This family of dimeric nucleases was the first to be discovered and has been used for gene targeting experiments for the past 20 years. They usually recognize extended DNA sequences (14-44 bp) that are very rare in genomes⁷¹. Given the length of their recognition site, and contrary to bacterial restriction enzymes, meganucleases can tolerate few mismatches in their target sequence even though the cleavage efficiency tends to decrease⁷². Meganucleases are generally coded within introns or inteins, although freestanding members also are present⁷³. To the present

knowledge their sole purpose is to promote the mobility within the genome of the DNA sequences from which they are coded, so they could be classified as selfish genetic elements⁷², even though recent studies indicate that the complex nature of the interconnection of the nuclease with the host may suggest a mutualistic relationship⁷⁴. The mechanism used to colonize cognate alleles lacking the endonuclease coding sequence involves the generation of a DSB or a DNA nick in a genomic locus homologous to the parent site (from here the alternative name of homing endonucleases, HE), followed by repair by homologous recombination and gene conversion that disrupts the target site to avoid further invasion⁷⁵. Five main families of HE have been identified according to sequence and structure: GIY-YIG, HNH, His-Cys box, PD-(D/E)XK and LAGLIDADG, which has been found in all the main taxonomic groups and is the most well studied family⁷². Thanks to the length of the recognition sequence of most HE, differently to type II bacterial restriction endonucleases, they can be exploited for many molecular biology applications where rare or single cuts are needed (for examples of recognition sites see **Fig. 4**). Nowadays most applications are based on a restricted number of well-studied natural nucleases⁷⁶ and even though the number of HE still to be discovered and characterized remains prospectively high⁷², the targetable sequences are still too few to allow enough targeting diversity in complex eukaryotic genomes. To overcome this issue, the modification of the natural recognition sites of the I-SceI and I-CreI nucleases was obtained by altering specific amino acids making contacts with the target DNA⁷⁷⁻⁸¹. Another strategy is based on the combination of different DNA binding subdomains obtained from different HE in order to generate chimeras with an hybrid specificity inherited from each of the two parental nucleases^{82,83}. In addition, the targeting range of chimeric nucleases can be further tuned using mutagenesis on each of the different subdomain (**Fig. 4**)⁸⁴. Other approaches involving high-throughput generation of HE variants to explore a larger space of target sites have been devised^{84,85}. Given the extensive mechanistic and structural knowledge of the active site of some of these enzymes, it has been demonstrated that their cleavage activity can be modulated in order to obtain nicks in the target DNA instead of double strand breaks^{86,87}. Homing endonucleases have been extensively used *in vivo* in several model organisms⁸⁸⁻⁹³ to induce genomic DSB even though the more frequent targets were integrated reporters containing the natural site

recognized by the nuclease. There are few reported cases in which engineered nucleases have been used to introduce DSB in non-natural sequences containing disease-causing mutations to promote gene targeting and correct the defect; two examples are mutations in the XPC gene connected with xeroderma pigmentosum⁹⁴ and a mutation in the RAG1 gene causing SCID (Severe Combined ImmunoDeficiency)⁹⁵.

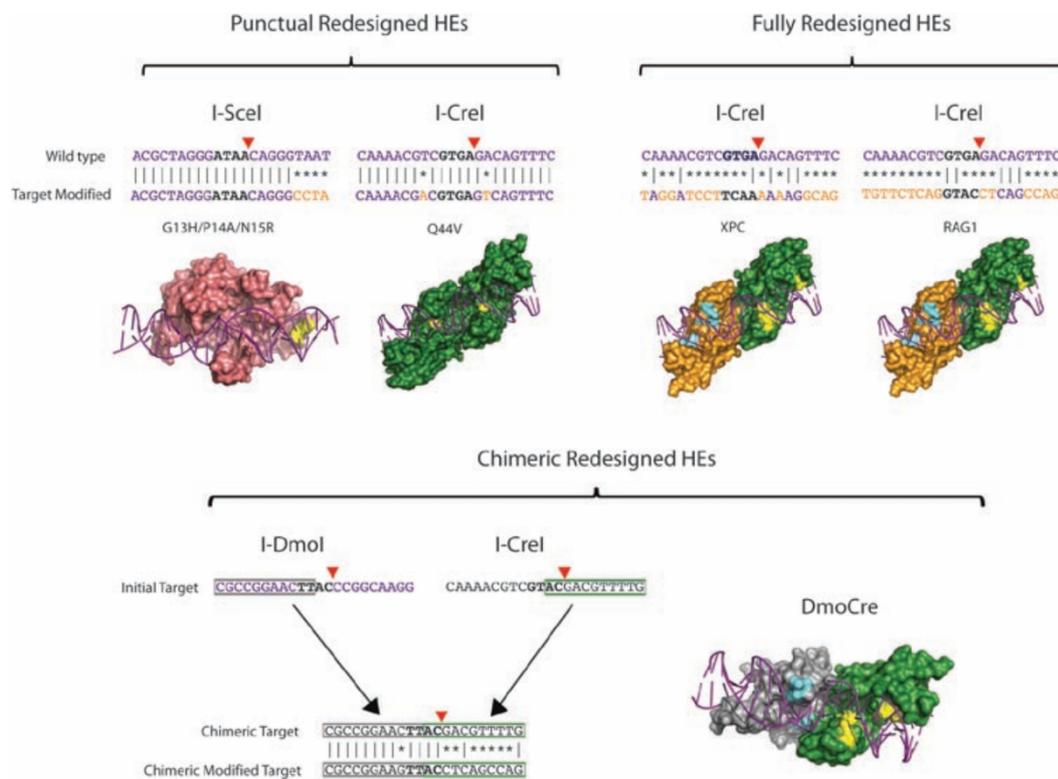


Figure 4. Engineering of homing endonucleases. New nuclease specificities can be obtained by engineering more or less extensively the meganuclease structure (upper panels). Alternatively, redesigned chimeric variants with hybrid target specificities can be obtained by combining the recognition domains of different natural occurring enzymes (lower panel). The mutations introduced in the structures are indicated in different colours. Adapted from Prieto et al., Crit. Rev. Biochem. Mol. Biol., 2012.

Zinc finger nucleases (ZFN). The necessity to obtain nucleases with suitable properties for the application in genome editing, such as the ability to specifically recognize unique genomic sequences and the amenability for retargeting towards different loci⁹⁶, led to the development of zinc finger nucleases. These have been obtained by linking a DNA binding domain commonly found in several eukaryotic transcription factors, called zinc finger proteins, to the nuclease domain of the FokI restriction enzyme⁹⁷. The nuclease portion of ZFNs is composed by the FokI catalytic

domain and needs to dimerize in order to generate a DSB at a specific target site (**Fig. 5**)⁹⁸. This feature, combined with the weak interaction between the two monomers, is crucial to ensure the specificity of ZFNs as only two adjacent zinc finger DNA binding domains will be able to recruit their FokI moieties close enough to favour dimerization and cleavage. Further engineering of the FokI domain allowed the generation of ZFNs that work as obligate heterodimers^{99,100}, thus avoiding the possible targeting of alternate sites by homodimers. Additionally, the modulation of the length of the linker peptide that separates the DNA binding domain and the catalytic portion of the nuclease allowed the modification of the spacing requirements between the two hemi-sites necessary for the cleavage of the target locus¹⁰¹, further increasing the flexibility of the system. The DNA binding domain of zinc finger nucleases is composed by a tandem array of Cys2-His2 motifs (fingers)¹⁰², each of them binding specifically to 3 bp (**Fig. 5**), with early nucleases being composed of three different fingers with the ability to recognize a 9 bp sequence. Since zinc fingers work as dimers the whole target sequence measures 18 bp, similarly to meganucleases. Later studies employed nucleases with up to six fingers per monomer, further improving the specificity of the approach⁹⁶. However, it must be underlined that longer recognition sequences not always necessarily produce an enhanced binding specificity as the tolerance for mismatches could also be increased. Structural analyses have revealed that individual fingers are able to bind DNA independently to neighbouring motifs¹⁰³, giving the possibility to generate zinc finger nucleases with custom specificity using modular assembly approaches¹⁰⁴. Using this strategy, new specificities are generated by the sequential assembly of different fingers previously isolated for their ability to bind to a specific DNA triplet^{105,106}. This method, however, oversimplifies the interaction dynamics inside the DNA binding module as it does not take into account that the contacts between different fingers or the interactions established by each finger with the target DNA outside its proximal triplet could have an effect on the zinc finger behaviour as a whole¹⁰⁷⁻¹⁰⁹. This prompted researches to find new ways to assemble ZFNs that explicitly take into account context-depend effects inside zinc fingers arrays: some examples are the oligomerized pool engineering (OPEN) system and its variants^{110,111}, where ZFNs are selected in bacteria by a two-step process, first singularly and then as an assembled array, or context-depend assembly (CoDA)¹¹². Both

methods, however, have some limitations in the range of sequences they can target. Another approach uses the modular assembly of two-fingers modules previously isolated for their ability to bind to a specific 6 bp sequence¹¹³. In this way, the number of untested module-module contacts are reduced to a minimum while maintaining the ease and speed of the modular approach, even though the initial screen to identify the specificity of double-finger modules is labour intensive.

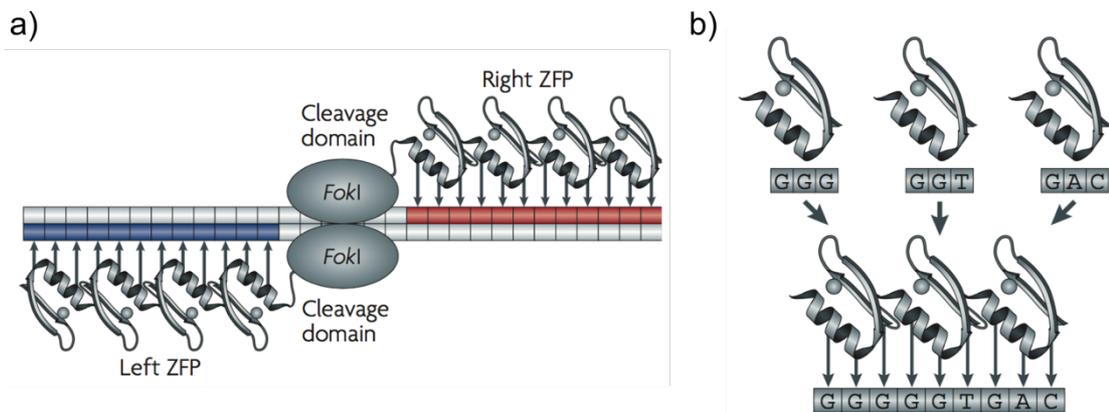


Figure 5. Zinc-finger nucleases. (a) Schematic representation of a zinc-finger nuclease dimer bound to its target DNA. Each monomer contains a FokI nuclease domain that upon dimerization cleaves the target sequence. A short spacer sequence (5-6 bp) separates the two recognition sites. **(b)** Zinc finger proteins are modularly assembled from finger monomers (in general 3-6 fingers arrays are used) each recognizing a 3 bp sequence. Adapted from Urnov et al., *Nat. Rev. Genet.*, 2010.

Zinc finger nucleases are among the first engineered endonucleases used in genome editing experiments, which involved the generation of knockout fruit flies¹¹⁴. Gene disruption experiments have then been performed in all main model organisms, with the consequent generation of transgenic lines^{111,115-117}, as well as in a plethora of other plants and animals¹¹⁸. The technology has been extended to mammalian cell lines¹¹⁹, implementing the possibility to generate simultaneously multiple knockouts¹²⁰ or large deletions⁶⁰. In addition, by exploiting the recombinogenic nature of ZFN-induced DSBs, it was possible to increase the efficiency of gene editing⁴² or gene addition⁵⁴ in different mammalian cellular models. As stated before, the possibility to dramatically enhance homology directed repair in cells through the introduction of targeted genomic lesions represents probably the major achievement in the genome engineering field and a turning point with previous editing approaches. The same technologies have been also employed to target specific loci in human stem cells¹²¹⁻¹²³, which are increasingly used as surrogate models for human tissues in basic research as well as in therapeutic and

drug discovery applications. The possibility to tailor custom ZFNs to target endogenous genomic loci with precision is very attractive for several clinical applications, from the correction of monogenic diseases to the employment for cancer therapy or against infectious pathogens¹²⁴. Different studies have shown the possibility to edit and correct disease-causing mutations involved in sickle cell anemia¹²⁵, α 1-antitrypsin disease¹²⁶ and in the alpha-synuclein gene associated with Parkinson's¹²⁷ in human induced pluripotent stem cells. In particular, ZFN-mediated gene disruption underwent a clinical trial for the treatment of glioblastoma through gene editing mediated disruption of the glucocorticoid receptor gene in T cells used for cancer immunotherapy (clinical trial number NCT01082926). Other clinical trials to treat HIV-1 infections by knocking-down the main viral co-receptor^{67,68,128}, the chemokine binding protein CCR5, have been completed (clinical trials number NCT00842634 and NCT01044654). CCR5 has been shown to be essential for viral propagation in the host but is dispensable for a correct development of immune cells, since ~1% of individuals is homozygous for a deletion of 32 bp in the CCR5 gene (Δ 32 mutation)⁷⁰. The treatment can be applied *ex vivo* both to T cells⁶⁸ and CD34⁺ hematopoietic stem cells⁶⁹, that can be re-infused into the patient to give rise to an HIV-resistant immune system. Proof of principle of the concept has been indirectly obtained by CCR5 Δ 32 homozygous bone marrow transplantation into an HIV-positive leukemic patient led to the disappearance of the virus from the blood stream¹²⁹. Overall, if compared to homing endonucleases, the flexibility of design of ZFNs allows the specific targeting of a wide range of genomic loci and makes them a good candidate for therapeutic applications, as demonstrated by their use in several clinical trials, as well as a major improvement in basic research. However, there are still some drawbacks in this technology: the complexity connected to the design and the assembly of the zinc finger protein module required to bind the target DNA as well as the lack of absolute specificity, that causes the introduction on unwanted cleavages into the target genome¹³⁰, are issues that can limit the broad applicability of ZFNs.

Transcription activator-like effector (TALE) nucleases (TALENs). To address the limitations connected to the ZFNs technology, a new class of guided nucleases was recently presented and rapidly developed: the TALENs. The basic design of TALENs is

similar to the one of ZFNs, where a DNA binding domain that recognizes a specific sequence is fused to an unspecific endonuclease domain obtained from the FokI enzyme (**Fig. 6**). As with ZFNs, also for TALENs it is possible to introduce mutations in the nuclease domain in order to avoid homodimerization and reduce possible off-target cleavages and consequent toxicity¹³¹. The DNA binding domain is composed of highly conserved repeats isolated from Transcription Activator-Like Effector (TALE) proteins secreted in host plant cells through a type III secretion system by the bacterium *Xanthomonas* to alter gene transcription and favour infection¹³². TALEs bind DNA through a highly conserved array of short (ca. 30 amino acids, except the last repeat that is usually shorter) motifs flanked by additional domains at both ends, with each repeat recognizing a single nucleotide. Complete modularity differentiates this system from zinc fingers binding domains, in which context-dependent effects must be taken in consideration, making TALE design extremely easy and straightforward. In addition, virtually any desired sequence starting with a thymidine residue can be targeted¹³³, even though this requirement is not strict^{134,135}. In general, each TALEN is engineered to bind a 15-30 bp target site, conferring the ability to recognize a final 30-60 bp sequence to the cleavage-competent dimer. Several groups over time have optimized in different ways the TALE scaffold to eliminate non-essential portions of the protein from the two constant N- and C-terminal domains, consequently modifying the length of the spacer DNA that must separate the TALEN couple to ensure efficient cleavage¹³⁴⁻¹³⁹. Subsequent structural studies revealed an essential role in DNA binding, but without any contribution to specificity, for the region of the N-terminal domain immediately preceding the repeats. For this reason the domain must be retained in TALEN constructs¹⁴⁰. In addition, mutations have been inserted in the N- and C-terminal domains of TALENs to improve the efficiency of genome editing⁴⁴. As stated before, DNA recognition specificity is determined by the ability of each TALE repeat to bind to a specific nucleobase using two variable amino acids in position 12-13 of the repeat¹⁴¹. The experimental evidence that the length of recognition site correlates with the number of repeats contained in the TALE protein coupled with bioinformatics analysis of naturally occurring binding sites, allowed the determination of the code that governs the interaction^{133,141}. Crystal structures have revealed that TALE binds DNA as a right-handed superhelix, with each repeat forming a two-helix bundle containing a variable

loop that inserts in the major groove of the target DNA (**Fig. 6**). Residue 12 (histidine or asparagine) does not contact DNA directly but stabilizes the loop by hydrogen-bonding with the alanine in position 8. On the other hand, residue 13 makes contact with the corresponding base of the sense strand of target DNA^{142,143}. The majority of studies published so far employs just four different domains containing the hypervariable residues NN, NI, HD and NG to recognize guanine, adenine, cytosine and thymidine, respectively¹¹⁸. However, the NN di-residue is able to interact also with adenine, leading to unwanted binding to similar sequences. Other di-residues (NH and NK) have been identified for their ability to uniquely interact with guanine residues, restoring binding specificity. Nevertheless, the employment of the NK di-residue is responsible for a drop in TALENs cleavage efficiency that cannot always be tolerated^{135,144,145}. In addition, following the observation that in some instances methylated DNA was not cleaved by TALENs¹⁴⁶, different groups have shown that the NG di-residue or a truncated loop missing the amino acid in position 13 can bind to 5-methyl-cytosine, commonly found in CpG islands of mammalian genomes^{147,148}.

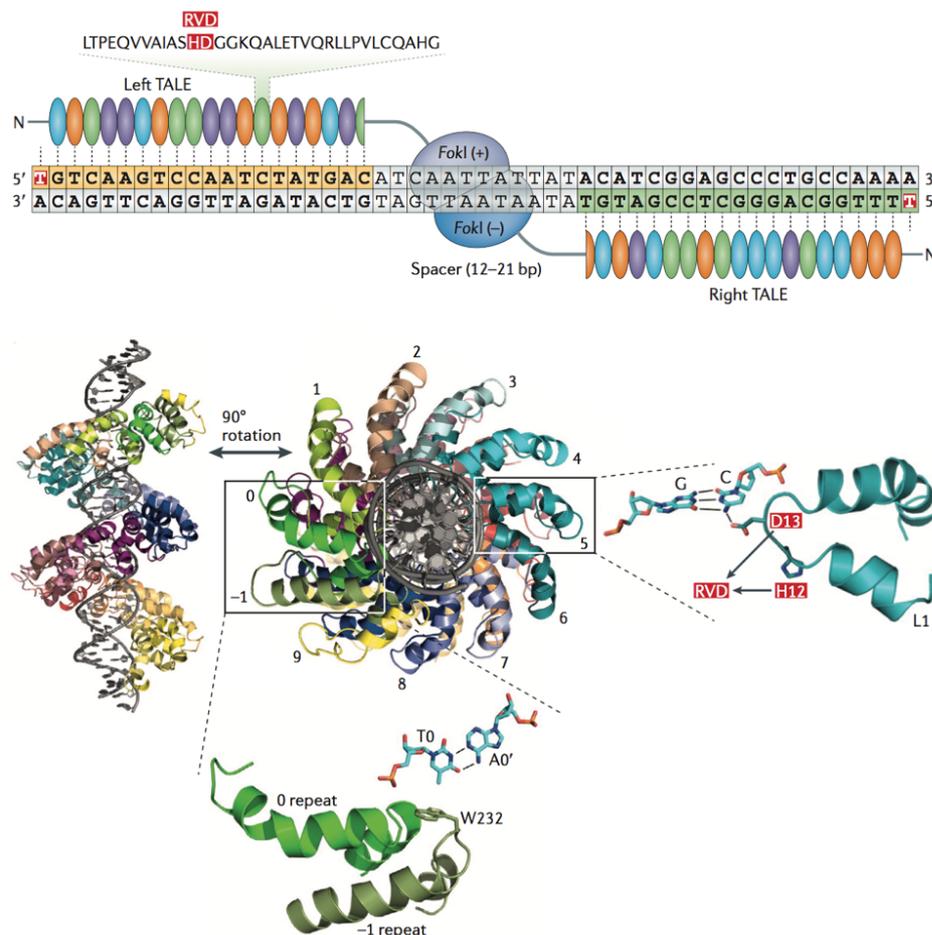


Figure 6 (on previous page). TALE nucleases. Schematic representation of a TALEN dimer bound to its target DNA. The FokI nuclease domain dimerizes to cut the target sequence. Each TALE monomer is composed of a 33-35 amino acid repeat that recognizes a single base and is assembled modularly to generate the final DNA binding domain, that interacts with a 15-30 bp target site. Base recognition employs a Repeat Variable Di-residue (RVD, in red) that contacts the target DNA major groove. The N-terminal repeats (0 and -1) contact an invariable 5'-thymidine.

The assembly of TALE repeats can be a challenging process and different methods have been proposed to minimize the effort: the restriction enzyme and ligation (REAL) method and its variants¹⁴⁹⁻¹⁵¹ are the more classical ones and involve the hierarchical assembly of increasingly complex arrays of TALE repeats by standard cloning techniques. A direct improvement of the latter approach is Golden Gate assembly, that is now widely employed in the field and makes use of type IIS restriction endonucleases that cut outside their recognition sequence and can generate any desired 5'-overhang sequence^{136,152-158}. By flanking each TALE domain with two type IIS restriction sites and by carefully designing the overhangs it is possible to assemble a TALE array (6-10 repeats) in a single ligation reaction. Higher order repeats fused with the constant portions of the TALEN can then be obtained by adopting the same strategy on previously built smaller arrays. Another approach uses ligation independent cloning (LIC) that is based on the very specific annealing of fragments long overhangs during transformation without the need of a ligation step or selection of single colonies on agar plates¹⁵⁹. In addition, all these processes could be sped up considerably by using pre-assembled libraries of multimers of TALE repeats. Other high-throughput approaches mainly based on solid-phase ligation strategies have been devised for industrial scale synthesis¹⁶⁰⁻¹⁶².

TALENs application to both basic and biomedical research builds on the previous widespread use of ZFNs for similar scopes, improving on the simplicity and reliability of the tool. In particular, different groups have shown that TALEN demonstrate higher specificity and less toxicity into cellular and animal applications, if compared to zinc finger nucleases^{46,131,138}. The application of the TALEN technology to laboratory animals led to the development of several knockout lines that could be used to model human diseases¹¹⁸. Similarly, genome editing experiments have been carried out in cell culture mainly by inducing knockouts by NHEJ-mediated repair^{134,135,138,152,156,161} but also exploiting HDR and double-stranded DNA templates to introduce specific changes into

the genome of clinically relevant cell types such as human pluripotent stem cells^{135,163}. Aside from generating DSB at target sites other applications have been designed for TALE proteins by fusion with domains different from the *FokI* nuclease. TALE-based transcriptional activators have been tested in plants^{144,154,158,164-166} and human cells¹⁴⁶ to modulate gene expression, obtaining at best a 2-30 fold increase in gene or protein expression, with the vast majority of TALE-based transcription factors showing only low levels of activity. On the opposite, by fusing TALE repeats to transcriptional repressors it was possible to downregulate gene expression in a sequence-specific manner in different model organisms^{145,165,167,168}. A recent study employed TALE proteins to transiently direct the binding of a set of three different epigenetic modifiers to specific genomic sites in order to obtain stable and inheritable silencing¹⁶⁹. Another fusion partner that has been tested in combination with TALEs is the catalytic domain of the DNA invertase Gin to obtain a targeted recombinase (TALER)¹⁷⁰. In principle, other fusion proteins can be generated to act on the cellular genome in a sequence specific fashion. Overall, TALE proteins improve the technology launched with zinc finger nucleases by increasing the specificity of DNA recognition and lowering the effort in designing each locus-specific binder.

The CRISPR/Cas system

A new class of programmable nucleases emerged in the last years, the Clustered Regularly Interspaced Palindromic Repeats (CRISPR)/ CRISPR associated (Cas) system, also known as RNA guided nucleases (RGNs), finding quickly broad application in genome engineering approaches and often superseding previous classes of editing nucleases. Its extreme technical simplicity and low cost made RGNs very popular between non-specialist scientists, transforming CRISPR-mediated genome editing into a mass-phenomenon.

The origins of RGNs sit into a completely separate scientific field, born when researchers begun investigating the functions of CRISPRs in bacterial genomes, almost 20 years after their discovery in *E. coli*¹⁷¹. CRISPR-Cas loci are present in all Archea and in almost half bacterial genomes, are constituted of a characteristic pattern of repeated

stretches of DNA alternated by spacer sequences and are usually located near *cas* genes (**Fig. 7**)¹⁷².

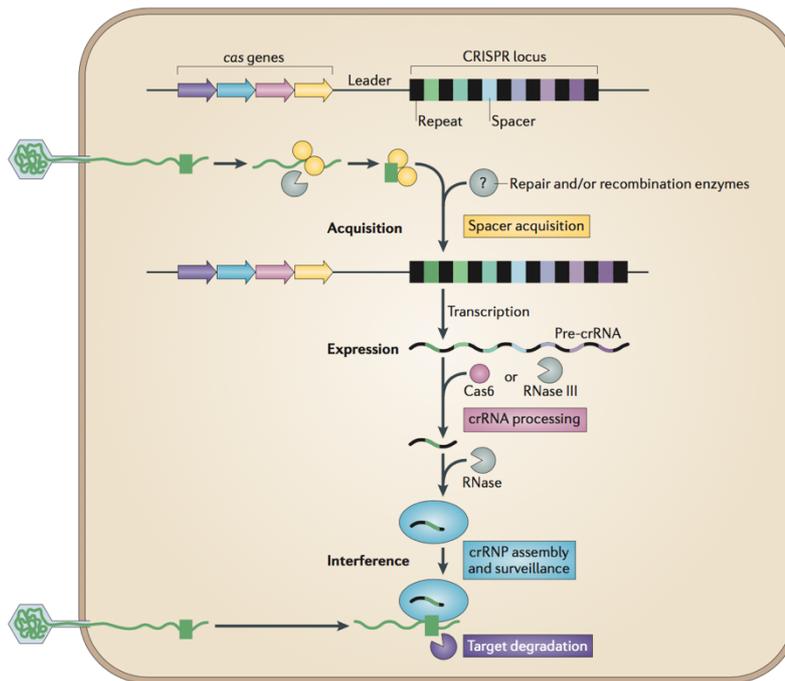


Figure 7. The bacterial CRISPR system. Bacterial adaptive immunity is mediated by CRISPR RNAs in complex with *cas* proteins. Following invasion by a foreign nucleic acid, new spacers are acquired and integrated into the CRISPR locus for later expression, processing and incorporation into crRNP for the interference stage. Adapted from van der Oost et al., *Nat. Rev. Micro.*, 2014.

Different studies determined that the spacer sequences found in CRISPR loci were identical to sequences present in the genome of bacteriophages and other mobile genetic elements¹⁷³⁻¹⁷⁵ to which the bacterium was resistant. This, combined with the observation that *cas* genes code for proteins with putative helicase and nuclease activities^{173,176}, led to the identification of the CRISPR system as a form of nucleic acid-based bacterial adaptive immunity¹⁷⁷. Experimental evidences indicated that this sort of adaptive immune system is based on ribonucleoprotein complexes (crRNP) composed of Cas proteins that cleave invading nucleic acids guided by RNAs transcribed from the spacer sequences found in bacterial genomes (crRNAs), exploiting base-pairing with the exogenous nucleic acid (**Fig. 7**)^{178,179}. The CRISPR/Cas loci are highly diverse due to the rapid evolution induced by the arms race with invading pathogens and, according to the present knowledge, are classified in two classes in accordance to the number of Cas proteins present in the complex that degrades invading nucleic acids with a further division in six types with many different subtypes¹⁸⁰. Each main CRISPR type is characterized by a peculiar Cas protein: type I systems contain Cas3, type II systems signature protein is Cas9, all type III systems are

associated with Cas10, type V systems are characterized by Cpf1 and type VI systems are identified by the C2c2 protein. Consequently, each CRISPR type is also characterized by a particular crRNP: the Cascade (CRISPR-associated complex for antiviral defence) for type I-A to I-F, type II systems use Cas9 complexes, while type III-A and type III-B are identified by Csm and Cmr crRNPs, respectively¹⁸¹. Despite the apparent complexity, Cas proteins can be divided in four main functional groups: nucleases and recombinases, involved in spacer acquisition; ribonucleases involved in guide RNA processing; proteins that together with guide RNAs form crRNPs; nucleases that cut target DNAs or RNAs. The first step in CRISPR/Cas mediated immunity involves the acquisition of new spacer sequences to be integrated into the host chromosome (**Fig. 7**). This process is spatially ordered and new spacers are always added upstream of the CRISPR locus^{173,177} after sampling by the conserved Cas1 and Cas2 proteins¹⁸² and processing to a specific spacer size¹⁸³. Some studies suggested that other Cas proteins, as well as some housekeeping proteins involved in DNA repair and recombination, are involved in spacer acquisition^{177,184-186}. The distinction between self and non-self in type I and II systems is obtained through short stretches of nucleotides collectively called protospacer adjacent motif (PAM) that are absent from the CRISPR spacer located into the bacterial genome but are present in the foreign nucleic acid^{187,188}. Several studies demonstrated that the presence of a PAM is essential for CRISPR/Cas mediated restriction¹⁸⁹⁻¹⁹⁴ as well as for the incorporation of new spacers¹⁸⁷. Type III systems use a PAM-independent self/non-self discrimination mechanism based on the pairing of the crRNA with specific sequences in the genome¹⁹⁵. CRISPR-arrays are transcribed as long RNA precursors (pre-crRNA) that are processed to generate mature crRNAs. In both type I and type III systems Cas6 is generally involved in this maturation step, generating a hairpin structure at the 3'-end of the RNA¹⁹⁶. In type I-E and type I-F systems Cas6 remains associated with the mature crRNA and becomes part of the crRNP after crRNA incorporation¹⁹⁷. In addition, the 3'-end of crRNAs is further trimmed after incorporation into crRNP¹⁹⁸. In type V systems, the maturation of crRNAs requires only the Cpf1¹⁹⁹ protein.

crRNPs differ in protein composition between different CRISPR types, even though effector complexes from type I and type III systems share several structural similarities (**Fig. 8**)²⁰⁰⁻²⁰². The prototype type I Cascade complex, belonging to type I-E systems, is

composed of a core of Cas5-Cas6-Cas7 proteins organized around the crRNA, with the addition of two other subunits (Cse1 and Cse2) with uneven stoichiometry^{178,203}. Other type I subtypes differ in the identity of the Cas protein involved in crRNP assembly even though the overall structure of the complex remains very similar, with a characteristic helical backbone^{204,205}. Similarly, type III complexes are composed of a different set of homologous Cas proteins, but nevertheless share several common structural features with type I RNPs^{200,202}.

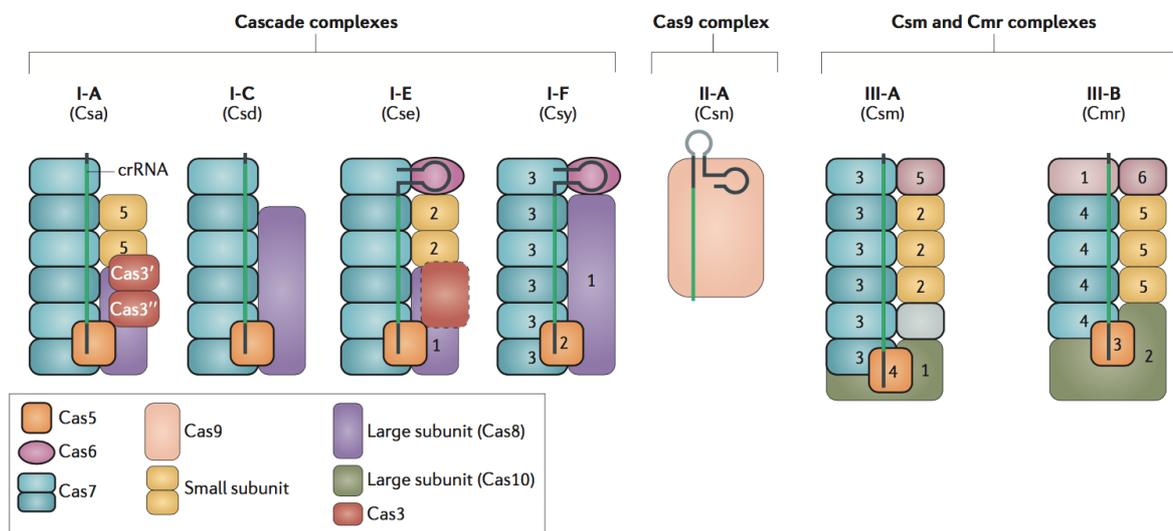


Figure 8. CRISPR effector complexes. Schematic representation of the subunit composition of the main three CRISPR-Cas types. Similar colours indicate homology between the different components of each complex. crRNA is represented in each complex, Cas9 is depicted bound to a synthetic single guide RNA. Numbers indicate the name of the subunit in the context of each particular CRISPR type. Cas3' and Cas3'', the dashed red subunit in the Cse complex and the grey subunit in the Csm complex are not well characterized. Adapted from van der Oost et al., Nat. Rev. Micro., 2014.

Target recognition and restriction proceed in a stepwise manner and begin with PAM recognition in the invading nucleic acid followed by DNA duplex melting and pairing between the crRNA and the target (sequence), starting from a seed sequence (7-8 nucleotides) proximal to the PAM^{192,206,207}. This induces a conformational change in the crRNP, activating its endonuclease function (either directly or *via* an additional subunit) and promoting target cleavage^{202,203,208}. In type I systems, the Cse1 subunit of the Cascade complex scans the DNA for PAM recognition^{193,203,206}, crRNA-DNA pairing is aided by the Cse2 subunit²⁰⁸ and a conformational change in the complex recruits the helicase-nuclease Cas3 that mediates complete degradation of the target DNA¹⁹⁴. The mechanistic details of type III-A systems are still to be elucidated completely, however,

it is known that Csm complexes are able to recognize foreign DNA in a PAM-independent manner¹⁹⁵ and that target degradation is mediated by a helicase-nuclease recruited after binding, possibly Csm6²⁰⁹. Type III-B systems are peculiar in their ability to target RNA instead of DNA²¹⁰. Crm complexes are in fact able to cleave target RNAs multiple times following a regularly spaced pattern²⁰⁰.

CRISPR type II systems: the Cas9 endonuclease

Different studies in *Streptococcus thermophilus* have shown that Cas9 is responsible for the defence against invading nucleic acids through the introduction of DSB into plasmids and phages^{177,211-213}. The CRISPR type II locus, that so far has been found only in bacteria, is characterized by the presence of an atypical family of small non coding trans-activating RNAs (tracrRNAs). TracrRNAs share a 24-nucleotide complementarity with the repeat region of crRNAs precursors and are responsible for directing the activity of the host protein RNase III, that in presence of Cas9 promotes the maturation of pre-crRNA into mature crRNA in which the spacer sequence is trimmed to a final length of 20 nucleotides. In the process tracrRNAs are also cleaved to their mature form of 75 nucleotides²¹¹.

Both mature crRNA and tracrRNA are essential to direct Cas9 cleavage towards a DNA molecule that contains a target complementary to the crRNA followed by an appropriate PAM sequence which is species-specific¹⁸⁹. The essentiality of the PAM sequence was demonstrated by interaction studies that revealed that Cas9 preferentially binds to PAM-containing DNA sequences while forming only transient interactions with other sequences^{214,215}. After PAM binding, Cas9 interrogates the flanking DNA for complementarity with the crRNA to initiate R-loop formation at the 3'-end of the guide RNA with sequential unwinding of the target DNA through a Brownian ratchet motion model (**Fig. 9**). Thus, R-loop association rates are influenced by the presence/absence of a PAM, while R-loop stability depends on the pairing between guide RNA and target DNA²¹⁴. It has been shown that perfect matching of a 12 nucleotide PAM-proximal seed region is important for proper target recognition and cleavage, while more PAM-distal positions can tolerate mismatches^{189,190,215}. Cas9 contains a HNH- and a RuvC-nuclease domain that are responsible for the cleavage of

the target DNA strand complementary and non-complementary to the crRNA, respectively (**Fig. 10**)^{189,190}. Cleavage products are characterized by blunt ends produced 3 nucleotides upstream of the PAM sequence²¹³.

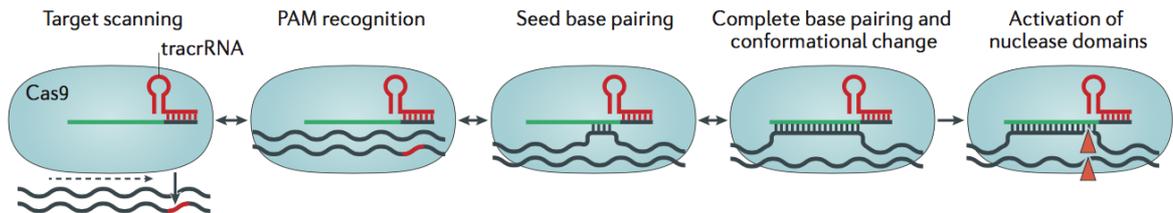


Figure 9. Cas9 target recognition and cleavage. Schematic representation of Cas9 mode of action: Cas9 in complex with both crRNA and tracrRNA scans the DNA for PAM sequences. PAM recognition is followed by seed pairing that upon perfect match is extended to the rest of the spacer sequence triggering conformational changes and R-loop formation. The two target DNA strand are then cleaved by distinct Cas9 nuclease domains. Adapted from van der Oost et al., Nat. Rev. Micro., 2014.

Cas9 proteins form a large family across the bacterial kingdom with members characterized by little sequence similarity and spanning different sizes, even though two main populations centred around 1100 and 1350 amino acids can be identified. Three principal sub-families have been proposed based on the architecture and the organization of each CRISPR locus: type II-A and type II-C, that comprise most of Cas9 proteins, plus type II-B²¹⁶. Structural analyses of prototype members of sub-families II-A and II-C, Cas9 from *Streptococcus pyogenes*, *Staphylococcus aureus* and *Actinomyces naeslundii*, revealed some common structural features and the general domain organization of this protein family²¹⁷⁻²¹⁹. Before binding to their target both proteins fold in a bi-lobed structure in which is possible to discriminate a nuclease lobe, that contains the conserved catalytic structural core, and an alpha-helical lobe that, together with the carboxy-terminal domain, are more diverse between different members of this family and are probably involved in the recognition of molecular partners peculiar to each Cas9 protein, such as guide RNAs and PAMs²²⁰. The transition from the *apo* state to the guide RNA-bound state corresponds to a significant structural rearrangement, involving in particular the alpha-helical lobe, with the formation of a central channel between the two lobes to home the target DNA-guide RNA duplex (**Fig. 10**)^{217,218}.

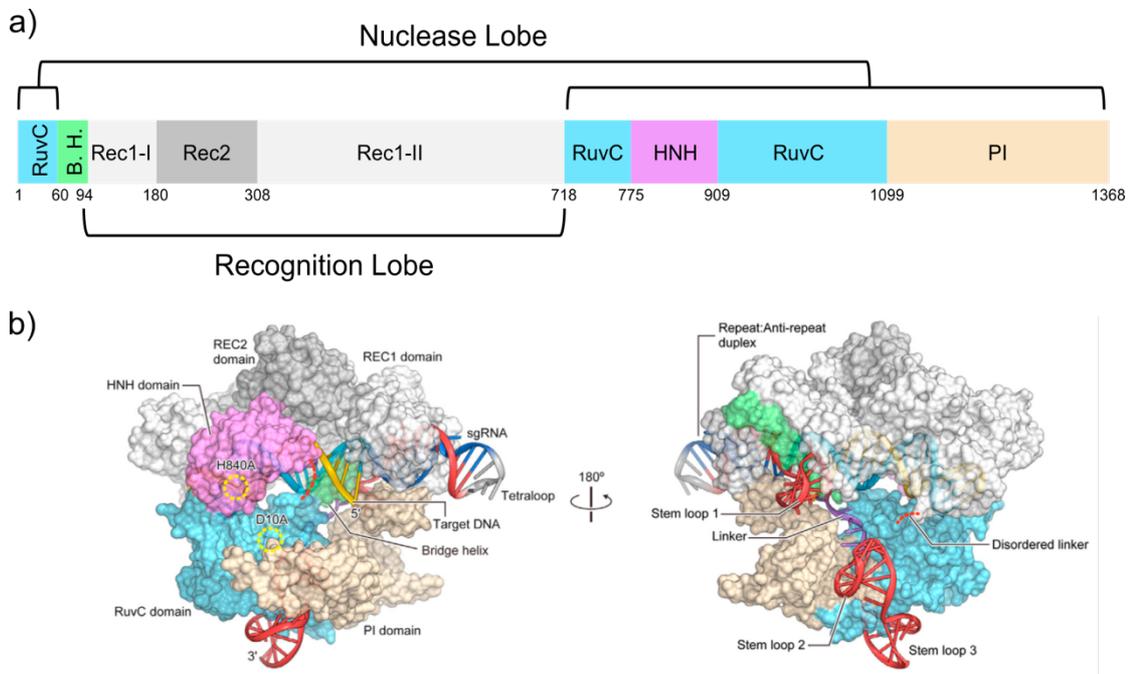


Figure 10. Cas9 structure. (a) Schematic representation of Cas9 domain organization. (b) Surface representation of Cas9 in complex with a single guide RNA and target DNA. A bilobed structural organization can be appreciated, with the alpha helical lobe and the nucleases lobe that fold to generate a central cavity to host the sgRNA:DNA heteroduplex. Colours in (a) correspond to surface colourings in (b). BH: bridge helix. Adapted from Nishimasu et al., Cell, 2014.

Cas9 interacts with the RNA-DNA heteroduplex independently of its sequence, recognizing the geometry of the complex rather than its base composition. On the other hand, the interaction with the guide RNA is sequence specific, explaining the strict preferences in crRNA/tracrRNA loading of each Cas9 orthologue. In addition, a conserved arginine-rich helix present in the N-terminal portion of the protein is essential for both RNA and DNA recognition and act as a hinge, connecting the two Cas9 lobes. The structural basis of PAM requirement and recognition has been elucidated²²¹, demonstrating that Cas9 makes specific contacts with the GG dinucleotide of the non-target DNA strand (residues R1333 and R1335 of *S. pyogenes* Cas9), without forming any interactions with the cognate bases on the target strand. A general model for Cas9 target recognition and cleavage has been proposed: in the absence of its RNA component Cas9 assumes an auto-inhibited conformation that is switched to an active conformation upon guide RNA binding; after this step Cas9 can interrogate putative target DNAs for PAM sequences. Upon PAM binding and R-loop formation, progressive RNA-DNA heteroduplex pairing unwinds the target DNA from the PAM-proximal end and generates an amenable substrate for Cas9-mediated

cleavage²²⁰. This general model, however, is still unable to explain the exact mechanism that leads to the cleavage of each target DNA strand, since structural studies generally employed Cas9 mutants in which at least one of the two catalytic sites is inactivated. Additionally, it is not clear how target DNA unwinding proceeds or if it is connected to the conformational changes induced by PAM binding. Recently, new insights on Cas9 cleavage mechanism have been obtained²²², showing that the conformational state of the HNH nuclease domain determines the extent to which Cas9 endonucleolytic activity is engaged, identifying an allosteric communication between the HNH and the RuvC domains to allow a concerted cleavage reaction.

Biotechnological applications of CRISPR nucleases

Among different RGNs, the type II system is particularly interesting for its dependence on a single multi-domain and multi-functional Cas protein (Cas9) and two short RNAs for its effector roles. In 2012 two different groups analysed the cleavage requirements of Cas9 from *Streptococcus pyogenes*¹⁸⁹ (SpCas9) and *Streptococcus thermophilus*¹⁹⁰ (St3Cas9) and demonstrated the possibility to redirect their target specificity by modifying the crRNA. Immediately after, the application of *S. pyogenes* Cas9 to edit the genome of eukaryotic cell lines has been reported by different groups that introduced indel mutations in different loci by exploiting NHEJ-mediated repair of Cas9-induced DSBs²²³⁻²²⁵. In one instance, a significant increase in the frequency of homologous recombination at the targeted site has been reported²²⁴. These results opened a brand new era in the field of programmable nucleases and, more broadly, in genome engineering. If compared to previous tools, such as ZFNs or TALENs, RGNs are characterized by a similar efficiency in targeting specific loci, combined with a much more straightforward design process for retargeting²²⁶. The generation of a chimeric molecule, the single-guide RNA (sgRNA), from the fusion between the crRNA and the tracrRNA through a four base pair loop¹⁸⁹ allowed to further simplify the experimental design, facilitating the delivery of all the RGN components (**Fig. 11**). In addition, the fact that Cas9 is not directly coupled with its sgRNA allows easy multiplexing experiments in which different guide RNAs are delivered in the cell at the same time to obtain mutations in different genomic loci²²⁷⁻²²⁹ or deletions and inversions⁶¹. This is a unique

advantage of the RGN platform relative to previous programmable nucleases technologies.

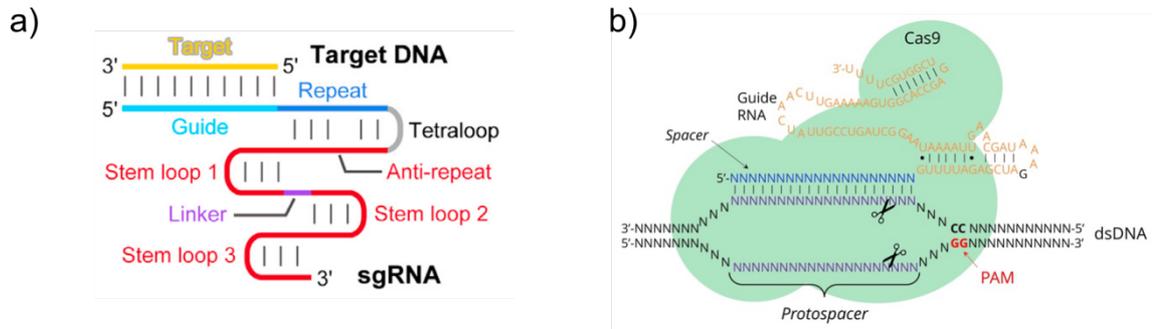


Figure 11. Single guide RNAs. (a) Schematics of the single guide RNA (sgRNA) engineered by the addition of a tetraloop between the crRNA and the tracrRNA. The stem-loop structure that characterizes the sgRNA is indicated in the scheme. (b) Cas9 in complex with a single guide RNA and the target DNA. (a) adapted from Nishimasu et al., Cell, 2014.

In the last years, Cas9-mediated genome editing has been efficiently applied to many different human and non-human cells and embryonic stem cells²³⁰⁻²³². Relevant applications are, for example, the generation of cancer models bearing genomic modifications similar to the ones found in patients' samples²³³⁻²³⁶ or proof-of-principle studies demonstrating the correction of genetic mutations responsible for inheritable diseases^{43,237,238}. The application of the Cas9 toolkit for the generation of animal models proceeded at a rapid pace and several transgenic species have now been obtained including, in addition to mice, worms, flies, fish, rats, rabbits, goats, sheep, dogs, pigs, and monkeys^{231,239}. This demonstrates the possibility of a further expansion of the pool of species amenable for genetic modification, leading to the generation of more efficient and reliable models for biomedical research. In addition, the possibility to directly inject Cas9 RNPs into fertilized zygotes can speed-up the process of gene modification in animals, bypassing the necessity of generating transgenic embryonic stem cells lines, shortening the time required to obtain a transgenic mouse or rat from a year to several weeks^{227,228,240,241}. In a similar way, Cas9-based genome engineering has found application also in plants, both in species used as research models and in crop plants like rice, wheat, sorghum, tobacco, sweet orange and liverwort²⁴², where in some instances high editing efficiencies have been observed, combined with the stable transmission of the alterations to the next generations²⁴³. Of note, the regulatory implications of such modifications are still not completely clear and will have to be

addressed in the next years²⁴⁴.

An immediate extension of the CRISPR/Cas9 technology is the high-throughput screening of genomic functions. The simple principle on which Cas9 retargeting is based allows the bioinformatics design and the preparation of chemically synthesized oligonucleotides libraries targeting hundreds of thousands specific sequences in the genome. These sgRNAs libraries are generally delivered as pools into cells using lentiviral or retroviral vectors, ensuring that each individual cell receives only one single sgRNA. Cultures can then be both positively or negatively selected for a specific phenotype and the connected sgRNAs are then identified by deep-sequencing and bioinformatics analysis²⁴⁵⁻²⁴⁸. Of note, a study reported the *in vivo* screen for genes involved in metastases formation and late-stage primary tumour development using cells manipulated *ex vivo* with whole genome sgRNA libraries²⁴⁹. Further implementations of this technology, allowed by the efficiency of the lentiviral delivery system, are the application of CRISPR screens to primary cells²⁵⁰, to obtain insights on physiologically relevant biological systems, or the design of multiplex screens in which more than one gene is targeted at time²⁵¹. Different studies have compared the performance of CRISPR screens with that of classic RNAi approaches concluding that RGNs outperformed previous experimental designs, with less variation in the results and fewer off-target effects, and suggesting that the combination of the data obtained from the two approaches could increase the overall performance^{252,253}.

Beyond its employment as a targeted nuclease, Cas9 has been modified in several ways to bring different biological functions to specific sites of the genome (for a summary see **Fig. 12**). The knowledge of the precise configuration of its two catalytic sites allowed the generation of catalytically inactive Cas9 mutants (dCas9, containing the mutations D10A and H840A)¹⁸⁹ that are able to bind the target DNA but cannot induce any DSB, becoming RNA-guided DNA-binding proteins. By fusing transcriptional repressors to dCas9, such as the Kruppel-associated box (KRAB) domain, the downregulation of the expression of protein-coding genes, as well as non-coding sequences including miRNAs and large intergenic non-coding RNA (lincRNAs), has been obtained in mammalian cells²⁵⁴⁻²⁵⁷. On the other hand, fusions with multiple copies of the VP16 transcriptional activator domain (VP64 or VP128) induced only mild upregulations of endogenous genes^{254,258,259}. For this reason, different strategies have been devised to increase the

efficiency of the activation process, observing a clear dependence on the genomic context of the gene²⁶⁰. Examples are the use of epitope arrays fused to dCas9 to allow the binding of several intracellular antibodies fused to a transactivator domain²⁶¹, the introduction of multiple MS2 hairpins in the sgRNA sequence to allow the recruitment of transcriptional activators through their fusion with the MS2-binding protein²⁶² or the engineering of stronger tripartite transcriptional activators composed of VP64, the activation domain of p65 and the Epstein-Barr virus R transactivator²⁶³. In addition, the design of more complex scaffold RNA (scRNAs) by introducing different hairpins into sgRNA in order to recruit endogenous regulators of transcription has been proposed²⁶⁴.

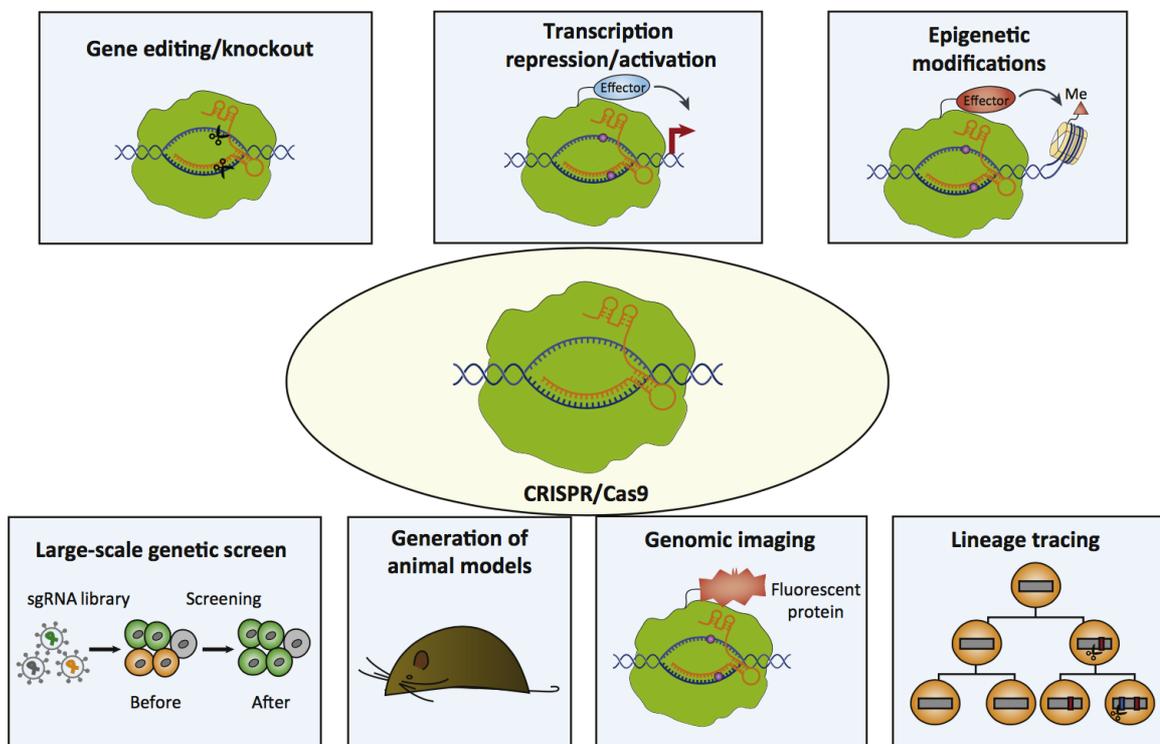


Figure 12. Biotechnological applications of Cas9. Several applications have been designed for Cas9, both as a targeted nuclease for genome engineering and as a programmable DNA binding domain to direct the desired biological function to specific genomic regions. Adapted from Wang&Qi, Trends Cell Biol., 2016.

Starting from similar concepts, dCas9-mediated epigenome engineering has been attempted by generating different fusions of dCas9 with epigenetic-modifying enzymes. An example is represented by the generation of a chimera with the catalytic core domain of the human histone-acetyltransferase p300 that is able to upregulate gene expression through acetylation of lysine 27 of histone H3 when bound to

proximal or distal enhancers²⁶⁵. Other tested chimeric molecules include fusions with lysine demethylase 1²⁶⁶ and different DNA methyltransferases^{169,267}. However, more in depth studies are needed to expand the epigenetic editing toolkit, to assess its complete orthogonality with endogenous mechanisms and to understand the fate of the modifications introduced, even though a recent investigation analysed the outcome of epigenome editing after long term culture¹⁶⁹.

Given its ability to target specific sequences, dCas9 has been employed to directly observe the organization and the dynamics of genomic sites of interest in cells. The first proof-of-principle study exploited a dCas9-eGFP fusion to track non repetitive and repetitive genomic loci such as telomeres^{268,269}. Building on this idea, by using different Cas9 orthologues tagged with alternative fluorescent proteins, the simultaneous tracking of different repetitive loci has been achieved²⁷⁰. A direct improvement of the latter system uses an array of epitopes (SunTag) fused to dCas9 that recruits multiple copies of a fluorescently tagged intrabody in order to amplify the signal produced by a single non-repetitive genomic locus²⁶¹. The development of Cas9-mediated fluorescence in situ hybridization (CASFISH) allows the visualization of both repetitive and non-repetitive loci in fixed cells and primary tissue sections using fluorescently labelled *in vitro* assembled dCas9 RNPs²⁷¹. This method represents a fast and convenient way to label endogenous loci without the need to denature DNA, thus preserving the original nuclear architecture and the spatial relationships between different genomic sequences and avoiding the artefacts typical of standard FISH techniques. dCas9 has additionally been used to image endogenous RNA molecules in living cells by using a sgRNA targeting a specific RNA combined with a stabilized PAMmer oligonucleotide that base-pairs with the same target and contains the PAM sequence necessary for dCas9 binding. This method allowed to track RNA export from the nucleus and the accumulation of specific mRNAs into stress granules following oxidative stress²⁷².

A recently reported unconventional application for Cas9 involves the editing of single bases within specific sequences without the induction of DSBs^{273,274}. These base editors are constituted by the fusion of a catalytically inactive Cas9 to a specific cytidine deaminase enzyme, that uses the single stranded DNA generated by Cas9 binding as a substrate for deamination, targeting different cytidine residues within the protospacer

sequence, according to the length of the linker between the two fusion partners. This base editing system can be improved by using a Cas9 nickase that, by introducing of a nick into the non-edited strand, enhances editing activity through the re-synthesis of the non-edited strand using the modified one as a template, thus incorporating the newly added mutation. Furthermore, the fusion to the base editor of a protein inhibitor of uracil DNA glycosylase (UDG) is able to boost editing efficiencies, blocking the engagement of the base-excision repair pathway^{273,274}. This approach can be used for the correction of pathogenic single base mutations without the risk of inducing DSB or indels and with a superior efficiency with respect to HDR-based protocols.

Limitations in the CRISPR/Cas9 technology

In the past years a plethora of new applications and techniques exploiting *S. pyogenes* Cas9 has been developed and even though this CRISPR nuclease has proven a flexible and powerful tool to modify the genome or study its functions, some drawbacks still remain. One issue is connected to the relatively big size of the Cas9 protein (~1400 amino acids, ~4,3 kilobases) that limits its delivery with some conventional vector systems, such as AAV vectors, that are commonly used in gene therapy approaches, but can ferry into the cell inserts up to 4,5 kilobases. Even though AAV-mediated Cas9 delivery in a single vector is technically feasible²⁷⁵, this leaves little room for flexibility in vector design. This explains the effort devolved by researchers into the identification of smaller Cas9 orthologues that could be more easily adapted to all kind of vector delivery systems, including AAV. Two smaller orthologues that have been isolated from *Neisseria meningitides* (NmCas9) and *Streptococcus thermophilus* strain LMD-9 CRISPR1 locus (St1Cas9) were tested for their ability to edit mammalian genomes^{223,276,277}. In addition, bioinformatics analyses aided the identification of an additional smaller Cas9 orthologue from *Staphylococcus aureus* (SaCas9) that was efficiently delivered *in vivo* using an AAV-based vector system to perform gene editing into liver tissue^{219,278}. An alternative approach is based on the reduction of SpCas9 dimension by eliminating non-essential portions of the protein. The deletion of the of the REC2 domain located in the alpha-helical lobe of SpCas9 has been reported, even though this shorter version of the nuclease shows reduced cleavage activity²¹⁸.

A second issue derives from the requirement of a specific PAM sequence in proximity of the intended genomic target. The canonical *S. pyogenes* Cas9 PAM corresponds to the trinucleotide 5'-NGG that should theoretically allow cleavage each 16 bp into any genome. However, the identification of an optimal sgRNA still remains not always possible, in particular when it is necessary to target specific nucleotides or short sequences and for homology-directed repair applications where the best efficiency is obtained only if cleavage occurs in close proximity to the desired alteration²⁷⁹. For this reason, the identification of new nucleases with different PAM specificities to expand the CRISPR toolbox can significantly increase the number of targetable genomic sites, with the possibility to choose between different Cas proteins according to each specific application. Example cited previously include *S. aureus* Cas9, characterized by the 5'-NNGRRN PAM sequence²⁷⁸, *N. meningitides* Cas9 that cleaves sites upstream the 5'-NNNNGATT PAM²⁷⁷ or *S. thermophilus* CRISPR1 Cas9 that recognizes 5'-NNAGAAW²⁷⁶. An additional Cas9 orthologue isolated from the CRISPR3 locus of *S. thermophilus* (St3Cas9) that cleaves sequences nearby the 5'-NGGNG PAM has been characterized¹⁹⁰ and tested for its ability to cleave mammalian genomes²⁷⁶. An alternative approach to address this limitation involves the generation of novel Cas9 variants with altered PAM specificities through protein engineering. Using a bacterial selection-based directed evolution screen, two Cas9 mutants with the ability to recognize 5'-NGA or 5'-NGC PAMs instead of the canonical 5'-NGG trinucleotide have been isolated²⁸⁰. In addition, the structural characterization of these variants was performed to understand the basis for differential PAM recognition and allowed the rational design of an additional mutant that specifically recognises the 5'-NAAG PAM²⁸¹. Similar protein engineering approaches have been used to broaden the targeting range of *S. aureus* Cas9 by relaxing its PAM recognition specificity to allow the cleavage of sequences followed by the NNNRRT PAM²⁸². In parallel, the need to understand the PAM requirements of newly discovered Cas proteins propelled the development of fast and reliable methods to identify and visualize PAM preferences²⁸³.

Other recently discovered editing systems

This continuous search for alternatives to SpCas9, in addition to the identification of several Cas9 orthologues, led to the discovery of other CRISPR-associated proteins with different properties. Recently, a new putative class 2 CRISPR system (type V) has been identified²⁸⁴ and structurally characterized^{285,286}. Its signature protein, the ~1300 amino acid Cpf1 protein, has been characterized as an RNA-guided nuclease that recognizes a T-rich PAM (TTTN-3') upstream of the target sequence and cleaves DNA in a PAM-distal position, generating staggered ends with 2-4 nucleotides overhangs¹⁹⁹. The latter characteristic could be particularly useful to introduce exogenous DNA sequences into the genome *via* a non-HDR mechanism²⁸⁷. In addition, the fact that DSBs are not introduced in proximity of the PAM sequence might allow a "second chance" cleavage mechanism where Cpf1 cuts again its target, tipping the balance between NHEJ and HDR in some experimental settings. In contrast to Cas9, Cpf1 contains a single identified RuvC-like nuclease domain and a second poorly characterized nuclease domain that together produce DSBs into the target DNA. The additional peculiarity of type V systems is that Cpf1-associated CRISPR arrays do not require a tracrRNA to be processed to their mature form, hence Cpf1 needs to incorporate only the crRNA, which is 42 nucleotides long with a spacer of 23-25 nucleotides, to be programmed for targeting specific sequences. Further analysis of several Cpf1 orthologues isolated from different bacterial strains allowed the identification of two enzymes (AsCpf1 and LbCpf1) that exhibit robust editing activity in mammalian cells¹⁹⁹. Furthermore, knockout mice have been generated by electroporation of Cpf1 RNPs into embryos²⁸⁸ or by co-injection of Cpf1 mRNA and crRNA into fertilized eggs²⁸⁹.

Another recently discovered member of Class 2 CRISPR systems is C2c2, that belongs to type VI systems and lacks homology with any other known DNA nuclease, but contains two Higher Eukaryotes and Prokaryotes Nucleotide-binding (HEPN) domains, generally associated with RNase activity²⁹⁰. Further characterization determined that C2c2 is indeed an RNase guided by a single crRNA composed of a 28 nucleotides spacer and a constant portion of the same length. No PAM is required for recognition and cleavage but the nucleotide located immediately downstream the target sequence, named Protospacer Flanking Site (PFS), must be different from G. C2c2 cleaves ssRNA

multiple times in a secondary structure-dependent manner and once activated is able to cleave unspecifically unrelated RNA molecules. It has been shown that C2c2 can downregulate mRNAs in bacteria, albeit causing appreciable growth restriction connected to promiscuous RNA cleavage. This effect could be exploited to induce growth arrest and programmed cell death in specific cells expressing a target RNA. Other biotechnological applications involving an enzymatically inactive form of C2c2 include RNA labelling and RNA targeting with specific effector modules, RNA re-localization to specific cellular sub-compartments or pull-down experiments to capture specific transcripts and their partners²⁹⁰.

Taking into consideration the continuous search for new methods to edit genomes, much excitement followed the discovery of a completely independent class of guided nucleases. A recent study²⁹¹ reported the editing of mammalian genomes using the Argonaute protein from *Natronobacterium gregoryi* (NgAgo). This is a DNA-guided nuclease that uses gDNAs of 24 nucleotides to generate sequence-specific DSBs without the need of a PAM sequence and with minimal off-target activity. Much debate followed the disclosure of these results, as many of the laboratories that tried to reproduce the data did not succeed in the attempt²⁹².

Safety issues and off-target activity

An important question connected to the use of programmable nucleases is whether their cleavage activity is limited to the intended site or if additional genomic loci are targeted unspecifically by the nuclease with the consequent induction of NHEJ-mediated indel mutations or gross chromosomal rearrangements, such as deletions, inversions or translocations, if multiple sites are targeted simultaneously²⁹³. Considering the not-so-far perspective of using designer nucleases for therapeutic applications, answering this question and addressing the related issue becomes essential since unwanted alterations could lead to unfavourable clinical outcomes. An interesting example in this respect comes from a therapeutic approach to treat HIV-1 infections by ZFN-mediated knockdown of the CCR5 viral co-receptor that is currently in a phase II clinical trial⁶⁷. Studies have shown that this ZFN cleaves also a highly homologous site in the CCR2 gene, possibly leading to chromosomal aberrations^{60,294},

and little is known on the effects that such off-target activity could have in patients. The same issues are also true for biomedical and basic research, where result interpretation requires an optimal knowledge of the experimental conditions and the eventual presence of confounding variables. It must be noted, however, that in the last case it is possible to devise different experimental strategies to control for nuclease off-target effects, similarly to what has been done in the past when using RNA interference protocols. The issue connected to off-target activity is common for all targeted nucleases and has been investigated in detail both for ZFNs²⁹⁵ and TALENs²⁹⁶. More recently, the discovery and application to genome editing of RGNs required further efforts to analyse the specificity patterns of these nucleases. The results of such studies will be presented more in detail in the next paragraphs.

Different groups reported the possibility that mismatches in the sgRNA-DNA pairing can be tolerated by Cas9, producing cleavages in sites that are only somewhat similar to the desired target (see **Fig. 13** for an example)^{189,223,297-299}. More extensive studies on mismatch tolerance by SpCas9 confirmed that, in general, sequences containing

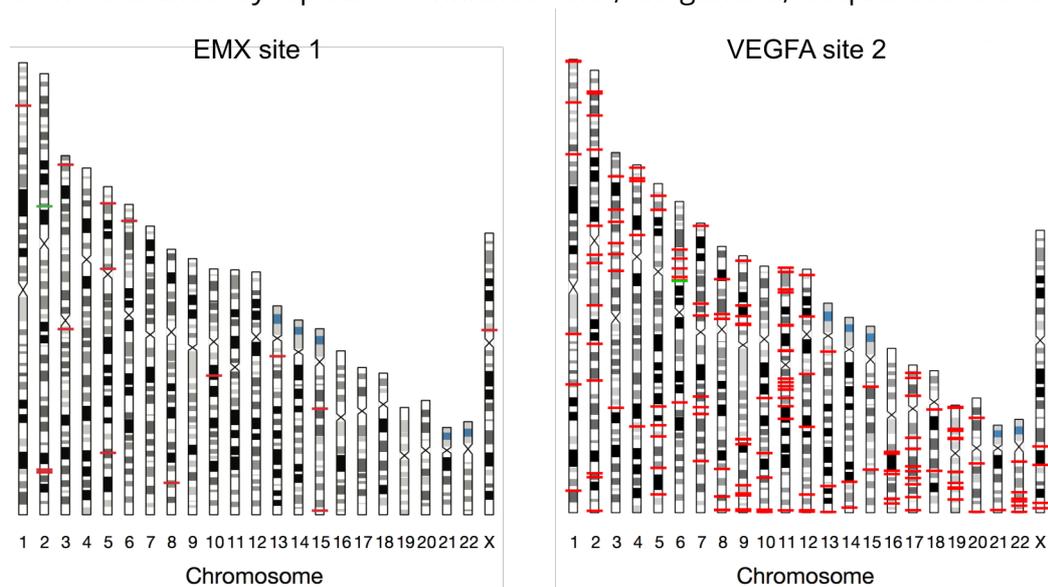


Figure 13. Cas9 off-target effects. GUIDE-seq analysis of genome-wide off-target sites for a conventional (EMX site 1) and a repetitive (VEGFA site 2) target. The position of off-target sites on chromosomes is indicated with red bars while the on-target is marked in green. Adapted from Tsai et al., Nat. Biotechnol., 2014.

mismatches in PAM-distal positions are targeted to a greater extent than sites with more PAM-proximal mismatches, but questioned the existence of a proper “seed region” where perfect match is always necessary to allow cleavage. Instead, a sequence-dependent and mismatch-dependent model has been proposed, with a clear relationship between number of mismatches and their relative positioning within

the guide sequence and SpCas9 cleavage efficiency^{300,301}. Another study used an *in vitro* selection protocol and purified SpCas9 RNPs to isolate from a randomized library sequences that are cleaved by selected sgRNAs. The subsequent analysis of the hits, some of which were further validated in mammalian cells, revealed a general trend for increased tolerance for mismatches in PAM-distal positions that is anyway dependent on the number of mismatches. Several exceptions to this general rule were observed and the working model in which only a so-called 10-12 nucleotide “seed sequence” is essential to specify the target sequence^{189,223} was refuted, stressing the importance of the pairing of all the 20 nucleotides of the spacer sequence to direct the nuclease. In addition, off-target sites identified with the *in vitro* screen and present in the human genome were tested for cleavage by expressing SpCas9 and the corresponding sgRNA in mammalian cells. Of all the tested loci only a minority was significantly modified³⁰², suggesting that other factors in addition to the actual target sequence, such as the accessibility of the genomic locus or the presence of epigenetic modifications, may determine the range of the off-target sites of a particular guide RNA. Other studies reported that RGNs can also cleave off-target sites where few extra or missing nucleotides are present, producing DNA or RNA bulges, respectively, to accommodate base-pairing of the remaining spacer portion³⁰³. In addition, genome-wide studies^{304,305} failed to identify a simple rule governing the relationship between the level of homology with the on-target sequence or the identity of the mismatched nucleotides and the level of cleavage of each off-target. These data clearly imply that the prediction of the specificity profile of different sgRNAs is neither simple nor straightforward. To aid the identification of putative off-target sites, that could be later experimentally tested, several software has been developed and is available online. These tools are generally based on the analysis of the sequence homology with the on-target site and take into account the number, position and nature of the mismatches to rank each off-target site.

Importantly, different studies identified a significant correlation between the amount of nuclease present in the cell or *in vitro* and the level of unspecific cleavages: increased amounts of SpCas9 were able to target sequences more dissimilar to the specific one³⁰⁰⁻³⁰². Similarly, the duration of nuclease exposure is another factor that can influence its off-target activity. The utilization of conventional plasmid transfection and

integrating viral vectors can lead to sustained expression of SpCas9 with the consequent induction of unwanted cleavages in a time-dependent manner, even though no particular cell toxicity has been observed so far³⁰⁶. Using whole genome sequencing at high coverage, two groups were able to identify only rare off-target mutations in human embryonic stem cells (hESC) and human induced pluripotent stem cells (iPS) clones^{307,308}, in contrast to what has been observed by others in established cell line models. This discrepancy could be due to the different experimental system used and to the fact that only clonal populations have been analysed in the stem cells models.

Several methods have been developed for the detection of both on-target and off-target modifications induced by guided nucleases. Sanger sequencing of DNA from individual clones is the most straightforward approach to measure the editing efficiency of known genomic loci (whether on- or off-target sites), but this method is suitable only for small sample sizes, it is expensive and time-consuming. High throughput sequencing is a direct evolution of the latter approach that allows the accurate measurement of indel frequencies of hundreds of on- and off-target sites in parallel, reaching a sensitivity that ranges from 0,001% to 1%, depending on the sequencing platform used. Other methods employ mismatch-sensitive nucleases such as the T7 endonuclease I (T7E1) or the CEL-I enzyme (Surveyor endonuclease). These enzymes recognize and cleave DNA heteroduplexes obtained by re-hybridization of amplicons relative to the locus of interest deriving from a population of edited cells. Since each sequence will either be wild-type or contain a variety of different mutations, upon annealing heteroduplexes will be formed. Agarose gel separation of the digestion products allows the estimation of the indel frequency at the tested locus, with a sensitivity around 1% and the possibility to detect even rarer events under optimal conditions²⁹³. In general, T7E1 is preferred to perform this assay since it is reported to be more sensitive to indel mutations than the CEL-I nuclease, while both enzymes can detect point mutations³⁰⁹. An alternative approach that has been recently developed is based on the bioinformatics analysis of Sanger sequencing chromatograms of amplicons relative to a target locus obtained from a population of edited cells³¹⁰. The Tracking of Indels by DEcomposition (TIDE) algorithm extracts and quantifies the information on all the indels present in a specific cell population from the capillary read

of the sequencing reaction, allowing a fast and cheap evaluation of the targeting efficiency with a sensitivity down to 1-2%.

All the previously presented methods require the precise knowledge of the location of RGN-generated indels. This is not always the case, in particular if we consider that some off-target sites could differ from their corresponding on-target by several nucleotides. A ChIP-Seq approach using dCas9 has been used to profile the DNA binding sites of the endonuclease in mouse embryonic stem cells and HEK293T cells revealing binding to thousands of sites that shared strong similarities with the on-target sites in the PAM-proximal region. However, when a selection of the identified sites was tested for cleavage with catalytically active SpCas9, the majority was found unmodified^{311,312}, indicating that DNA binding and cleavage are often uncoupled. Similarly, *in vitro* approaches using randomized oligonucleotide libraries can be used to characterize sites bound¹³⁵ or cleaved³⁰² by Cas9, however, also in this case limitations connected to the uncoupling between binding and cleavage, to the absence of most of the screened sites from the reference genome and to the impossibility to examine off-target sites that pair with the guide RNA through the formation of DNA- or RNA-bulges must be taken into consideration.

Recently, different methods have been developed to characterize the genome-wide *in vivo* profile of Cas9 cleavage sites (for a summary see **Fig. 14**). The use of Integrase Defective Lentiviral Vectors (IDLVs) capture has been introduced to track the formation of DSBs induced by ZFNs³¹³ and Cas9³⁰⁴. IDLV capture is based on the NHEJ-mediated integration of IDLVs at sites targeted by the nuclease, tagging these transient events that can be later identified through linear amplification mediated-PCR (LAM-PCR) and adapter ligation. Since the efficiency of IDLV integration is much lower than the mutation frequency at each particular site, the sensitivity of this method may not allow to identify all the *bona fide* off-target sites, even though increased detection can be obtained by upscaling the IDLV input dosage³⁰⁴. The overall sensitivity of this assay has been calculated to fall between 0,5% and 1%³⁰⁴. A direct improvement of this technique is the Genome-wide, Unbiased Identification of DSBs Enabled by Sequencing (GUIDE-Seq) where the integrase-defective lentiviral vector is substituted by a blunt-ended double-stranded phosphorothioate oligodeoxynucleotide that is captured at on- and off-target sites with higher efficiency to allow their isolation by PCR and high-

throughput sequencing, with a sensitivity down to 0,1% (see **Figs. 13** and **14**)³⁰⁵. The majority of the sites identified for each tested sgRNAs was further validated by targeted deep-sequencing, demonstrating that the captured sequences corresponded to *bona fide* off-targets³⁰⁵. High-throughput genomic translocation sequencing (HTGTS) and its improved version LAM-HTGTS exploit the translocations induced by RGN-generated DSBs at on- and off-target sites to detect unspecific cleavage events. HTGTS uses the on-target site as a “bait” to catch “prey” sequences that correspond to off-target cuts, that are then amplified by LAM-PCR and sequenced^{314,315}. Another approach captures DSBs induced after nuclease expression in fixed and permeabilized cells^{278,316}. Breaks Labelling, Enrichment on Streptavidin and next-generation Sequencing (BLESS) relies on the ligation of biotinylated oligonucleotides to DSBs after cell fixation to allow the identification of captured loci after DNA fragmentation through deep sequencing. Since this method identifies only the DSBs present at the time of fixation, it is probable that many off-target sites are missed, resulting in poor sensitivity, with the additional risk of incurring in artefacts due to the fixation procedure itself. It must be noted that, since they identify *in vivo* generated DSBs, GUIDE-Seq, HTGTS and BLESS, will also capture naturally occurring nuclease-independent DSBs. These hits, together with PCR artefacts generated during library production, constitute false-positive signals that must be discarded during the analysis of the sequencing data by filtering for sequences that are homologous to the on-target site. It has been calculated that up to 95% of captured sites are discarded during these filtering steps²⁹³. A last method to profile off-target activity relies on the *in vitro* digestion of cell-free genomic DNA. Digested genome sequencing (Digenome-seq) uses *in vitro* Cas9-digested genomic DNA to identify off-target cleavages by whole genome sequencing³¹⁷. Since digested fragments should present the same 5'-end given that no end-repair takes place, it is possible to look for straight alignments of sequence reads into the genome to identify cleaved sites, reaching a sensitivity below 0,1%, similar to the detection limit of targeted deep sequencing. Interestingly, the authors succeeded in validating by targeted deep-sequencing only a small group of captured sites³¹⁷.

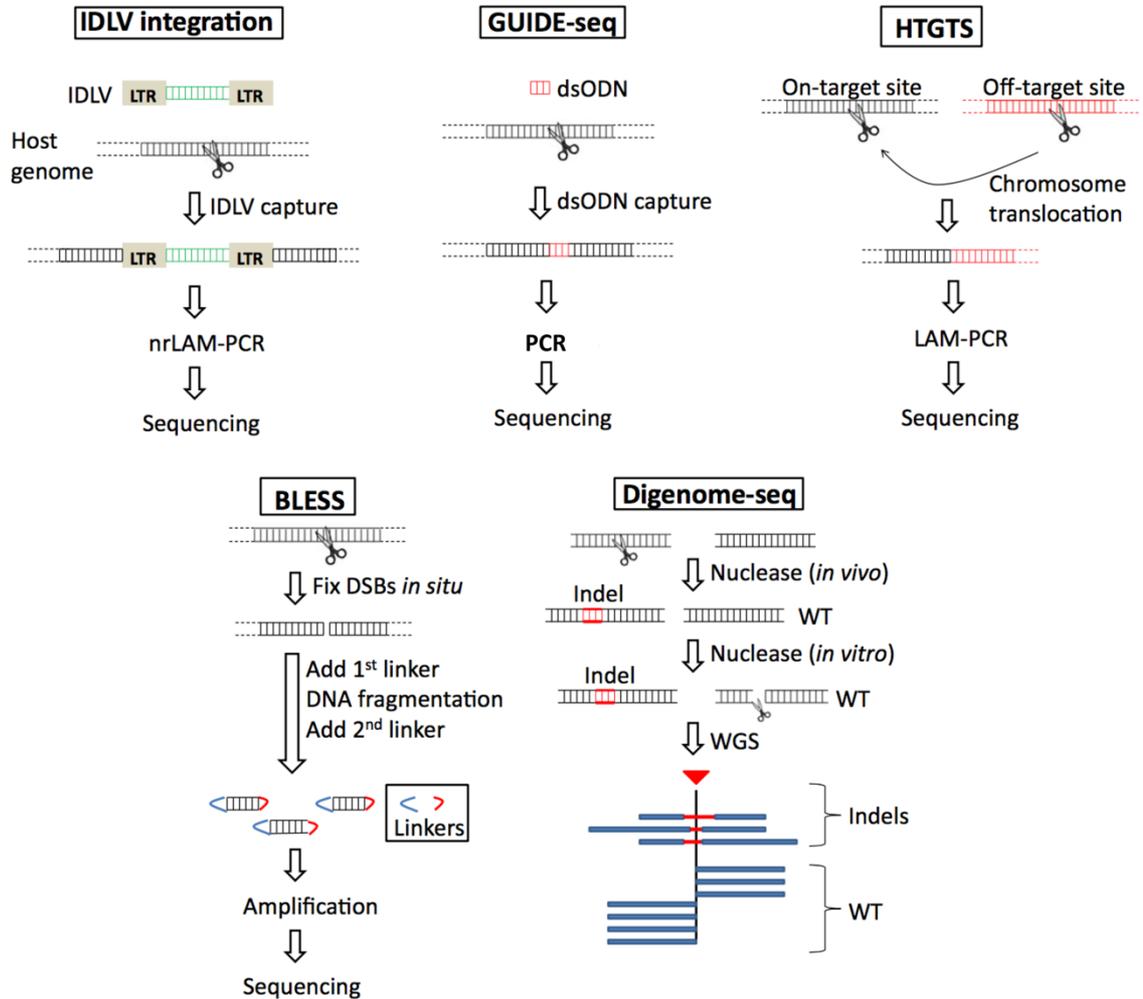


Figure 14. Genome-wide detection of Cas9 off-target activity. Schematic representation of the different unbiased methods that have been developed to investigate the genome-wide profile of Cas9 off-target activity. Adapted from Yee, FEBS J., 2016.

Even though it is not well known to what extent Cas9 cleavages are limited by chromatin organization, some studies investigated the effects of nucleosomal positioning on editing efficiency^{318,319} so it is possible to speculate that the set of off-target sites identified by *in vivo* capture will vary between cell lines, given the different chromatin organization and accessibility in each cell type. Digenome-seq, on the other hand, uses naked DNA as a substrate and for this reason is not sensitive to this kind of variations. In addition, GUIDE-Seq and HTGTS are dependent on NHEJ-mediated repair of nuclease-induced lesions that can introduce large insertions or deletions, making more difficult to pinpoint off-target sites²⁹³

A comprehensive comparison between HTGTS, GUIDE-Seq, Digenome-seq and BLESS has not been published yet. Only a single sgRNA has been profiled with the first three

methods showing that each technique was able to detect a different group of potential off-target sites indicating that no single approach is comprehensive. A small set of common loci, which had been previously validated, were identified by all three techniques³¹⁷.

Of note, these unbiased genome-wide detection technologies could be also applied with minimal modifications to the study of ZFNs and TALENs off-target activity.

Methods to reduce Cas9 off-target activity

The possibility that off-target genomic cuts may be generated during targeted nuclease-mediated genome editing represents the main technical limit for the development of therapeutic strategies to treat patients. Different approaches have been proposed to reduce unwanted mutations, demonstrating that a major need in the field is the generation of genome editing systems with no off-target activity. Interestingly, the application of unbiased methods to detect the genome-wide off-targets of the newly discovered type V Cpf1 RGN revealed that it has an intrinsic higher fidelity than SpCas9. In particular, GUIDE-Seq analysis of 20 different crRNAs was unable to detect any genomic off-target for more than half of the tested sites³²⁰ and similar results were obtained using the Digenome-seq method³²¹. These data suggest that, in perspective, the discovery and the application of new CRISPR nucleases with reduced off-target activity, such as Cpf1, could be a promising way to move towards an error-free genome editing toolbox.

As a starting point, when using all classes of targeted nucleases, a careful choice of the target sequences may help reducing off-target cleavages. For RGNs, the design of guide RNAs that differ from any other site in the genome by several nucleotides (at least 2-3) is an important way to reduce unspecific cuts. Indeed, genome-wide studies have demonstrated that it is possible to find gRNAs with no detectable off-target activity³⁰⁵. A wide range of off-target prediction tools, many of them working on different genomes, has been developed by different labs and is available online: Cas-OFFinder from the Kim lab, CHOPCHOP from the Church lab, CRISPR Design from the Zhang lab, CRISPR design tool from the Broad Institute, CRISPR/Cas9 gRNA finder from the Lin lab, CRISPRfinder from the Pourcel lab, ZiFiT Targeter from the Joung lab and E-CRISPR from

the Boutros lab. However, it must be noted that the set of predicted off-target sites on one hand is in general much bigger than the actual group of mistargeted loci identified by unbiased genome-wide methods, while on the other fails to include some of the sites that are captured by experimental methods^{305,317}. In general, however, choosing gRNAs that minimize the number of identified off-target sites, especially the ones with few mismatches with target sequences, could increase the chances of reducing unwanted damages to the genome.

A second approach is based on the modification of the sequence of guide RNAs. To optimize transcription from the RNA polymerase III U6 promoter *in vivo* and particularly for their *in vitro* production through T7 polymerase transcription, gRNA sequences should start with a G nucleotide in the former case and a GG dinucleotide in the latter. The configuration of the spacer of guide RNA molecules can thus correspond to G+N₂₀, G+N₁₉ with the first G matching to the target sequence, or GG+N₂₀, according to the different strategies. One study reported that the addition of two additional guanine nucleotides at the 5'-end of the sgRNA is able to abolish cleavage activity against the majority of validated off-target sites for four different guide RNAs, as measured by the T7E1 assay (**Fig. 15a**). Of note, a side-effect of this modification was a drop in on-target activity for two out of four tested sites²⁹⁸. The mechanism behind this effect is not yet clear and could involve guide RNA stability, concentration and structure. An additional explanation could be connected to an alteration of the interaction dynamics between the gRNA and the target DNA that leads to a decreased tolerance for mismatches, impairing non-specific cleavages. An additional strategy involves the truncation of sgRNAs (tru-gRNAs) in order to obtain a 17-18 nucleotide complementarity with the target sequence, instead of the classical 20 base-pairs (**Fig. 15b**)³²². Tru-sgRNAs are more sensitive to mismatches if compared to standard guides, owing to their reduced binding energy with the target DNA, and are able to reduce off-target cleavage up to several thousand-folds, reaching undetectable levels for some selected sites, without appreciable modifications in the on-targeting activity^{305,322}. However, it is still important to select tru-gRNAs with minimal homologies to the genome, since the reduction of the spacer length does not decrease the targeting range of the guide and possibly makes more difficult to avoid homologous sites containing only few mismatches with the intended target³²².

The precise knowledge of the architecture of SpCas9 active sites allowed the generation of single mutants (D10A and H840A) capable of cleaving only one of the two target DNA strands, generating a nick. The use of paired nickases (**Fig. 15c**) has been shown to reduce off-target activity by different groups: instead of a single RGN, two Cas9 nucleases are directed to opposite DNA strands by two separate sgRNA, allowing the formation of two nicks that combined can generate indel mutations or can favour HDR using single stranded DNA oligonucleotides as templates^{298,323,324,325}. Close proximity between the two nicks (between 4 and 100 bp) is important to obtain the desired effect, a concept similar to the one behind ZFNs nickases³²⁶, even though single Cas9 nickases can be used alone to stimulate HDR over NHEJ-mediated repair for safer gene modification approaches²³¹. The fact that different groups have shown that monomeric Cas9 nickases are indeed able to induce DSB into the genome on their own^{224,322-324} or induce base substitution at their target site with high efficiency³²⁷ raises concern over the necessity of having two separate gRNAs to target a single locus: even though the unspecific cleavages relative to one guide could be reduced, other new off-target sites could be generated by the second gRNA²³¹. In addition, the experimental design involving Cas9 nickases is not straightforward as the identification of correctly spaced and efficient pairs of guides is not always trivial. The generation of programmable nickases has been reported also TALE-based platforms³²⁸.

A strategy which is closely connected to the latter and further reduces off-target effects employs pairs of catalytically inactive Cas9 molecules fused through an appropriate linker to the FokI nuclease^{327,329}, in direct parallel with the strategy adopted with ZFNs and TALENs (**Fig. 15d**). In this particular setting, dCas9 acts as a DNA binding domain to bring in close proximity two FokI monomers to promote cleavage. Two specifically-oriented gRNAs separated by 15-25 bases are necessary to target each locus and the average reported cleavage efficiency is similar to that of paired Cas9 nickases, but lower than wild-type SpCas9. Importantly, since FokI monomers are not catalytically competent, mutagenic activity deriving from single dCas9-FokI fusions is avoided, contrary to previous approaches, while the strict topological requirements for gRNA positioning makes very unlikely the dimerization of FokI at off-target sites^{327,329}. This approach has been combined with tru-gRNA to further minimize the undesirable mutagenic effects of dCas9-FokI monomers³³⁰.

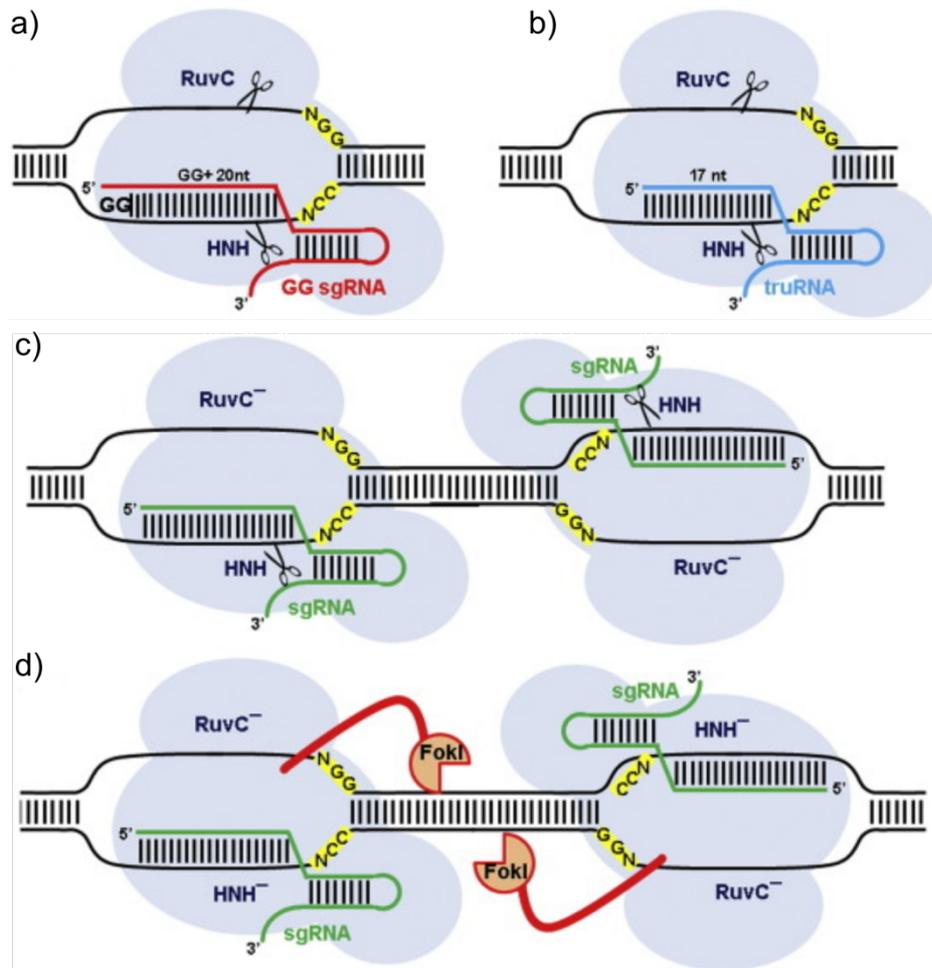


Figure 15. Methods to reduce Cas9 off-target activity. Different methods have been developed to reduce the risk of Cas9 off-target activity: **(a)** the addition of a 5'-GG dinucleotide to sgRNAs or the use of truncated guides (truRNAs) **(b)**; the employment of paired Cas9 nickases in which one of the two Cas9 catalytic domains has been inactivated **(c)**; the generation of catalytically inactive dCas9 proteins dependent from a FokI nuclease domain for target cleavage **(d)**. Adapted from Bortesi&Fischer, *Biotechnol. Adv.*, 2015.

The generation of chimeras between SpCas9 and programmable DNA binding domains (pDBDs) has been proposed as an alternative way to increase targeting specificity³³¹. In particular, the mutagenesis of key residues involved in PAM recognition to attenuate SpCas9 DNA binding affinity and broaden its PAM requirements allowed the generation of variants that were dependent on the attached pDBD for efficient function. The study suggests that the presence of the pDBD is able to increase the local concentration of Cas9, bypassing the kinetic barrier to R-loop formation that follows PAM recognition. Whole genome analyses to detect unspecific cleavages, combined with targeted deep sequencing of validated genomic off-targets, demonstrated that a

R1335K Cas9 variant fused to zinc finger proteins targeting a sequence proximal to the on-target locus was characterized by a significant increase in specificity³³¹.

Other important factors influencing the number of off-target modifications are the amount and persistence of SpCas9 expression in target cells: high concentrations of the nuclease are reported to increase off-site cleavage, whereas lowering the amount of SpCas9 increases specificity³⁰⁰⁻³⁰². Transient SpCas9 expression is indeed sufficient to permanently modify the target genomic locus with decreased off-target activity as demonstrated by the enhanced specificity obtained through direct delivery of recombinant SpCas9-sgRNA complexes into target cells through electroporation³³², lipofection³³³ or cell penetrating peptides³³⁴. Cas9-sgRNA RNP complexes induce mutations at the target site immediately after delivery and are quickly degraded by the cells afterwards, reducing the chances of off-target cleavages. In addition, it has been suggested that RNP delivery could be more efficient and less stressful in non-transformed cell types such as primary cells and pluripotent stem cells with the additional advantage of removing completely the possibility of unwanted integration into the genome of foreign nucleic acids coding for components of the genome editing platform²⁹³. Of note, the delivery through RNPs to reduce unwanted cleavages has been applied to other CRISPR nucleases, such as Cpf1^{288,321}. Similar approaches based on the delivery of mRNAs or proteins have been employed with ZFNs^{111,335,336} and TALENs^{151,337,338} in order to tightly control their intracellular levels and reduce off-target activity.

Transcriptional regulation of Cas9 expression offers an additional layer to control the amount and the duration of nuclease persistence into the cell^{339,340}. Alternatively, different studies have shown that it is possible to modulate Cas9 activity directly into the cell³⁴¹. Following the publication of detailed structural data on Cas9 that allowed the identification of two relatively independent protein lobes, different strategies have been developed to regulate the forced assembly of split variants of Cas9 in cells by using small molecule-regulated inteins³⁴², rapamycin-mediated interaction^{219,343} or exploiting photoinducible dimerizing protein domains³⁴⁴. The additional advantage of using split-Cas9 systems is the reduced size of its components that allows the packaging in low capacity viral vectors, such as AAV-based vectors³⁴⁵. An alternative photoactivatable Cas9 system relies on the incorporation of an unnatural photocaged

lysine into Cas9 to inactivate its catalytic sites. Irradiation with ultraviolet light allows the tight control of Cas9 activity, enabling the editing of the desired locus, albeit with lower efficiencies if compared to the wild-type protein³⁴⁶. An additional hurdle connected with this system is the necessity to modify cells to express an engineered tRNA synthetase to incorporate the desired unnatural amino acid into the nuclease. A different approach to address the specificity issue involves the modification of Cas9 amino acid sequence to isolate variants with increased discriminative power. During a screen for SpCas9 variants with altered PAM specificities, the D1335E mutation has been isolated for its ability to confer increased selectivity for the canonical NGG PAM as well as for its increased targeting specificity as determined by testing against a panel of 25 previously validated off-target sites relative to three different sgRNAs and using genome-wide profiling with the GUIDE-seq method²⁸⁰. Recently, two groups have reported the structure-guided rational engineering of SpCas9 variants characterized by a lower propensity to cleave off-target sites^{347,348}. The first study postulates that since Cas9 nuclease activity is stimulated by the separation of the two target DNA strands, by reducing its helicase activity target unwinding will be more dependent on gRNA:DNA pairing and mismatches between the sgRNA and the target sequence would be less tolerated, resulting in lower cleavage activity at off-target sites. The analysis of the crystal structure of SpCas9 in complex with guide RNA and the target DNA allowed the identification of a positively charged groove located between the HNH, RuvC and PI domains, likely to be involved in stabilizing the unwind target DNA. Variants containing alanine-substitutions of different charged residues within this groove were tested for their ability to reduce off-target activity. After a combinatorial screening two triple mutants (eSpCas9 v1.0 and v1.1 characterized by the substitutions K810A, R1003A, R1060A and K848A, R1003A, R1060A, respectively) were characterized by lower unspecific activity both at selected sites and according to unbiased genome-wide analyses³⁴⁸. A second study focused on the reduction of the unspecific interactions between Cas9 and the target DNA in order to eliminate the energy excess that characterizes the complex and allows the cleavage of mismatched sites, as previously suggested³²². Structural studies have identified different residues making contacts with the phosphate backbone of the target DNA strands and their mutation into alanines allowed the generation of a high-fidelity Cas9 variant (SpCas9-HF1, containing the

mutations N497A, R661A, Q695A, Q926A) with significantly reduced genome-wide unspecific activity, even though some off-targets relative to repetitive and conventional loci were still cleaved by SpCas9-HF1. Alternative variants have been generated by introducing additional mutations to SpCas9-HF1 further decreasing off-targeting activity, even though a parallel decrease in on-target cleavage efficiency has been observed³⁴⁷.

Altogether these efforts clearly demonstrate that a major need in the field is the generation of genome editing systems with no off-target activity.

Identification of high-fidelity Cas9 variants using a yeast-based screening

Rationale of the experiments

The number of biotechnological applications involving Cas9 has seen a huge increase in the past years, driven by the flexibility and the efficacy of this new genome editing tool. Cas9 has been applied with success in a variety of organisms, demonstrating the robustness that characterizes a game-changer technology. Importantly, both basic research and therapeutic-oriented studies, aside from efficacy, require specificity during editing procedures. However, several groups have shown that Cas9 cleavages into the genome are not always directed to the intended sites and unwanted lesions can be introduced in loci sharing different levels of similarity with the selected target. In addition, the prediction of such unwanted activity is difficult and often unreliable due to the absence of simple rules governing the phenomenon. Similarly, the assessment of the damages produced by nuclease expression is not always simple, and the results obtained with different techniques are often not in accordance. Hence, the development of methods to enhance the specificity of the CRISPR toolkit could represent a valid approach to eliminate the off-target issue at its roots.

Different strategies have been proposed to reduce the introduction of unwanted genomic mutations such as the tight control of SpCas9 intracellular levels, the introduction of engineered gRNAs characterized by shorter protospacers with less complementarity to the target sequence (tru-gRNAs), the fusion of SpCas9 to specific DNA-binding domains to direct its binding or the exploitation of paired SpCas9 nickases and paired catalytically inactive dCas9 fused to the FokI endonuclease domain. However, none of these approaches completely solves the issue due to drawbacks intrinsic to each technology, such as an increasing complexity that often leads to loss of cleavage efficiency and ease of use, coupled to incomplete off-target control. Recently, two groups have reported the structure-guided rational engineering of SpCas9 variants characterized by a lower propensity to cleave off-target sites.

The fact that SpCas9 can cut sequences that are slightly different from the intended target indicates that binding and cleavage are only partially dependent on the sgRNA-DNA pairing due to the extensive and unspecific contacts established between SpCas9 and the DNA phosphate backbone that are appreciably lowering the stringency of the target recognition step, as suggested by previous studies. Differently to previous reports, we wanted to identify amino acid substitutions leading to an error-free SpCas9 in an unbiased fashion through a random mutagenesis process in order to screen a wider mutational space, if compared to rational protein engineering approaches, and increase the likelihood of identifying better performing SpCas9 variants.

Methods

Plasmids and constructs. The plasmid p415-GalL-Cas9-CYC1t, designed by George Church's Lab, was used to express Cas9 in yeast (Addgene #43804)³⁴⁹. The Cas9 CDS was originally codon-optimized for applications in human cells^{224,349} and its expression is driven by the galactose inducible GalL promoter. The p415 plasmid contains the LEU2 gene as a yeast selectable marker and the CEN6/ARSH4 centromeric sequence that allows its propagation at low copy number inside yeast cells. To allow the precise removal of the Rec1-II domain by restriction digest, synonymous mutations were generated through PCR to introduce two restriction sites, NcoI and NheI, upstream and downstream of the Rec1-II domain, respectively (for primers, see Appendix Table 1). The expression cassette for the sgRNA was obtained from the p426-SNR52p-gRNA.CAN1.Y-SUP4t plasmid (Addgene #43803)³⁴⁹. In order to swap the original spacer sequence with the desired target, an assembly-PCR based strategy was adopted. The 5' portion of the sgRNA expression cassette was PCR-amplified using the T3 forward primer (annealing before the SNR52 promoter) and a reverse primer annealing immediately upstream of the spacer sequence and containing a 5' overhang corresponding to the desired on-target sequence (see Appendix Table 1). The same was done for the 3' fragment of the sgRNA, using the primer T7 reverse primer and a forward primer annealing immediately after the spacer sequence and containing a 5' overhang antiparallel to one previously mentioned. The assembly reaction to obtain the gRNA cassette was prepared by mixing both PCR amplicons and performing a single step of denaturation, annealing and extension, followed by an exponential amplification using only the T3 and T7 external primers. The resulting fragment was then gel purified and blunt-end cloned into pRS316, a low copy number centromeric plasmid carrying an URA3 yeast selectable marker, pre-digested with SacII/XhoI and blunted, generating the pRS316-SNR52p-gRNA.ON-SUP4t plasmid. We decided to swap the sgRNA plasmid backbone since we discovered that p426-based plasmids, due to their dependence on a 2 μ origin of replication that requires specific cellular machinery to be replicated, were not sufficiently stable in the yeast strains used during the experiments. If not correctly replicated, random segregation between progeny cells

during mitosis could dilute away the plasmids. On the other hand, pRS316 centromeric plasmids are always correctly segregated during mitosis.

For the expression of SpCas9 in mammalian cells we employed a pX330-U6-Chimeric_BB-CBh-hSpCas9 (Addgene # 42230)²²³ derived plasmid, where the sgRNA coding cassette has been removed, pX-Cas9. The SpCas9 coding sequence has been human codon optimized and its expression is regulated by a CBh promoter. In addition, two nuclear localization signals (NLS) have been added to the N- and C-terminus of SpCas9 to allow nuclear import and a triple FLAG is positioned at the N-terminal end of the protein to facilitate detection. The plasmids coding for improved Cas9 variants were obtained by sequential site directed mutagenesis starting from the pX-Cas9 plasmid. For the expression of previously published enhanced SpCas9 mutants the VP12 (Addgene #72247)³⁴⁷ and the eSpCas9(1.1) (Addgene #71814)³⁴⁸ plasmids were used. Desired spacer sequences were cloned as annealed oligonucleotides with appropriate overhangs into a double BbsI site located upstream the guide RNA constant portion in a pUC19 plasmid containing a U6 promoter-driven expression cassette. The same pUC19 plasmid containing an optimized guide RNA constant region²⁶⁸ was used for the preparation of optimized sgRNAs. For the experiments involving lentiviral vectors, the lentiCRISPRv1 transfer vector (Addgene #49535)²²³ was employed together with the pCMV-delta8.91 packaging vector and pMD2.G, coding for the vesicular stomatitis virus glycoprotein (VSVG), to produce viral particles. The lentiCRISPRv1 transfer vector contains an expression cassette for a codon-humanized version of a N-terminally FLAG-tagged SpCas9 fused through a 2A-peptide to the puromycin coding sequence to allow selection of transduced cells. A U6-driven expression cassette transcribes the sgRNA. Annealed oligos corresponding to the desired spacers were cloned into the guide RNA using a double BsmBI site. The lentiCRISPRv1-based vectors coding for enhanced SpCas9 variants were generated by swapping part of the SpCas9 coding sequence with a PCR fragment corresponding to the region of the CDS containing the mutations (for primers, see Appendix Table 1). A complete list of the guide RNA target sites is available in the Appendix.

Yeast culture. The yLFM-ICORE yeast strain^{350,351} was used to generate the reporter yeast strains used in this study. Synthetic minimal media (SD) were employed in all

yeast experiments (yeast nitrogen base without amino acids 6.7 g/L, L-isoleucine 600 mg/L, L-valine 150 mg/L, L-adenine 200 mg/L, L-arginine 20 mg/L, L-histidine 10 mg/L, L-leucine 100 mg/L, L-lysine 90 mg/L, L-methionine 20 mg/L, L-phenylalanine 50 mg/L, L-threonine 200 g/L, L-tryptophan 20 mg/L, L-uracil 20 mg/L, L-glutamic acid 100 mg/L, L-aspartic acid 200 g/L, L-serine 400 mg/L, D-(+)-glucose 20 g/L). Single amino acids were omitted according to the experimental setup, when selective medium was required. For the induction of Cas9 expression, 20 g/L D-(+)-galactose and 10 g/L D-(+)-raffinose were used instead of dextrose. Specific medium for ADE2 mutants colour screening was prepared using low adenine concentrations (5 mg/L). When non-selective medium was required, YPDA rich medium was employed (yeast extract 10 g/L, peptone 20 g/L, D-(+)-Glucose 20 g/L, L-adenine 200 mg/L). All solutions were prepared using ddH₂O, filter-sterilized and stored at 4°C. Solid media were prepared by autoclave sterilization, adding 20 g/L of agar to the solution. All chemicals to prepare yeast media were obtained from Sigma-Aldrich.

Yeast transformation. The day prior to transformation, approximately 1 mm³ of the desired yeast strain was inoculated in 5 mL of rich medium or selective synthetic medium and grew overnight at 30°C while shaking at 200 rpm. The next day 3-5 mL of the culture were inoculated in a total volume of 30 mL of the same medium and grew at 30°C shaking at 200 rpm for further 2-4 hours. Cells were then harvested by centrifugation at 2000xg for 2', washed in 30mL of ddH₂O, centrifuged again at 2000xg for 2' and resuspended in 10 mL of LiAc/TE 1X (lithium acetate 0,1 M and Tris 10 mM EDTA 1 mM, pH 7.5). The solution was centrifuged again at 2000xg for 2' and resuspended in a proper volume of LiAc/TE 1X (100 mg of yeast pellet in 500 µL). The transformation mix contained 500 ng of plasmid DNA, 5 µl of carrier salmon sperm DNA (approx. 1 µg) previously sheared by sonication and boiled at 100°C for 10', 50 µL of resuspended yeast culture and 300 µL of polyethylene glycol (PEG) 500 g/L with a molecular weight of ~36,500 (Sigma-Aldrich) diluted in LiAc/TE 1X. After vortexing, the transformation mix was placed for 30' at 30°C and then heat-shocked using a dry bath for 30' at 42°C. Cells were then centrifuged at 3000xg for 3', resuspended in 5 mL of the appropriate SD selective medium or directly plated on selective SD agarose plates and incubated at 30°C. For spontaneous reversion frequency evaluation, after

transformation with p415-GalL-Cas9-CYC1t cells were grown in selective medium for 24 hours. The concentration of cells was then evaluated by measuring the OD₆₀₀ and 1000 cells were plated on selective plates depleted of leucine (SDI) or 10⁶ cells were spread on plates further depleted of adenine (SDIa) or tryptophan (SDIt), to evaluate the number of revertants for each locus.

Yeast colony PCR. Colony PCRs were performed by resuspending approximately 1 mm³ of yeast colony in 49 µL of ddH₂O. 1 µL of lyticase (10000 U/mL, Sigma-Aldrich) was added to digest the cell wall and the suspension was then incubated at 30°C for 30'. The cells were pelleted, the supernatant was removed and the dry pellet was boiled for 10' at 100°C. The pellet was then resuspended in 50 µL of ddH₂O and 5 µL were used as a template in the PCR reaction, using the Phusion High-Fidelity DNA Polymerase (Thermo Scientific).

Recovery of plasmid DNA from yeast. In order to isolate the mutant Cas9 plasmids from yeast, single colonies were grown overnight at 30°C shaking at 200 rpm in 5 mL of SD medium without leucine (SDI), to select for the presence of the p415-GalL-Cas9-CYC1t plasmid, while relaxing the selection on the guide RNA-expressing plasmid to induce its dilution and loss. The next day cells were harvested by centrifugation for 5' at 5000xg and resuspended in 250 µL of buffer A1 (Nucleospin Plasmid, Macherey-Nagel) containing 0.1 mg/ml of RNase A. Cells were then mechanically lysed by adding 100 µL of acid-washed glass beads (Sigma-Aldrich) and by vortexing continuously for 5'. Plasmid DNA was then recovered from the supernatant using standard miniprep silica columns, following the manufacturer's instructions. DNA was eluted in 30 µL of 10 mM TrisHCl pH 8.5. The eluted DNA was treated with NcoI and NheI that can digest only the sgRNA-expressing vector, in order to avoid contaminations since these plasmids are also selectable through an ampicillin resistance. After digestion, 10 µL of the mix were transformed using chemically competent *E. coli*. The plasmids recovered were then digested to double check their identity and then Sanger sequenced to identify the mutations introduced in the Rec1-II domain.

Assembly of modified TRP1 and ADE2 genomic cassettes. The DNA cassettes used to engineer the ADE2 (ADE2-Off1, ADE2-Off2, ADE2-Off3 and ADE2-Off4) and TRP1

(TRP1-On) genomic loci were built using a similar strategy. Two different colony PCRs were performed to amplify the two halves of each wild-type locus separately. The first one employed a forward primer upstream of the gene CDS and a reverse overhang primer containing the on- or off1-4-target sequence followed by the KpnI or BamHI restriction sites, respectively (see Appendix Table 2). All reverse primers contained a stop codon before the on/off-target sequence to ensure truncation of the protein. The second half of the cassette was assembled using a reverse primer which anneals downstream the ADE2 and TRP1 coding sequences and a forward primer which anneals 100 bp before the reverse primer used to build the first half of the cassette. In this way, when the two parts were joined together, the final construct contained a 100bp long homology region upstream and downstream of the on-/off-target sequences. In addition, these forward primers contained the same restriction site present in the reverse primer of the corresponding first half of the cassette. The TRP1 and ADE2 fragments were assembled by ligating the two halves digested with KpnI or BamHI (New England BioLabs), accordingly. The products were separated on an agarose gel to remove homoligation-derived fragments. The final cassette was enriched by PCR using the most external primers and directly transformed in yeast.

Generation of yeast reporter strains. The *delitto perfetto* approach enables the genetic targeting of specific loci with the practicality of a general selection system through the exploitation of the homology directed repair mechanism that is particularly efficient in yeast³⁵². The first step consists in the insertion of a COUNTER selectable REporter I-SceI cassette (CORE-I-SceI) in the specific locus of interest. The cassette contains a recognition site for I-SceI, as well as the coding sequence for the endonuclease itself under the control of the galactose-inducible GAL1 promoter, the resistance gene kanMX4 (G418) and the counterselectable marker URA3 gene from *Kluyveromyces lactis* (KIURA3). The CORE-I-SceI cassette was amplified with primers containing specific overhangs for the ADE2 and TRP1 loci (see Appendix Table 2). Each locus was edited sequentially, following the same procedure, starting from the ADE2 locus. 500 ng of locus-specific CORE-I-SceI cassette was transformed in yeast and cells were plated on YPDA plates and incubated at 30°C overnight. The next day, colonies were replica-plated on YPDA media containing 200 µg/mL of G418 (Invivogen).

Resistant colonies were screened for successful cassette insertion into the desired locus by colony PCR using primers annealing to the genomic sequences flanking the integration site and to the cassette. The CORE-I-SceI cassette integrated within the targeted locus was then swapped with the final edited sequence (TRP1-On, ADE2-Off1, ADE2-Off2, ADE2-Off3 and ADE2-Off4), generating a total of four different yeast strains characterized by the same on-target sequence and four different off-targets. The appropriate intermediate yeast strain containing the target CORE-I-SceI cassette was inoculated overnight in 5mL of YPDA. The next day, before transformation, the inoculum was resuspended in 30mL of synthetic medium containing galactose and raffinose instead of dextrose (SRG). This step is essential to induce the transcription of the I-SceI endonuclease which cuts its target site located within the CORE cassette. DSBs increase the normal frequency of HR-driven repair events, favouring cassette-swap with the desired new sequence. After 4 hours incubation in SRG medium, yeast was transformed with 500ng of the HR template containing the desired sequence following the standard transformation protocol. Transformants were then plated on SD containing 60 mg/L of uracil and 1 g/l of 5-fluoroorotic acid (5-FOA) (Toronto Research Chemicals). 5-FOA, in the presence of orotidine 5'-phosphate decarboxylase (encoded by *KIURA3*), is converted in fluorouracil which is a potent thymidylate synthase inhibitor. 5-FOA-resistant colonies were then replica-plated on YPDA and YPDA supplemented with G418, to further select for the loss of the CORE cassette. By comparing the two replica plates it is possible to select G418-sensitive FOA-resistant colonies that correspond to positive clones. Colony PCRs, performed using genomic primers that anneal upstream and downstream of the entire genomic locus, were analysed by Sanger sequencing to confirm the sequence of the edited locus. The newly generated yeast strains containing the modified TRP1 and ADE2 loci were called yACMO-off1, yACMO-off2, yACMO-off3 and yACMO-off4, characterized by a selected on-target sequence in the TRP1 locus and four different off-target sequences in the ADE2 locus, each containing a single mismatch with respect to the on-target sequence in a position that is more PAM-proximal for off1 and more PAM-distal for off4 (see Appendix Table 3).

Yeast screening for SpCas9 mutants. The mutants' library was generated by error

prone PCR (epPCR) using the GeneMorph II kit (Agilent). Following the manufacturer's instructions, the initial amount of template DNA (p415-GalL-Cas9-CYC1t) and the number of cycles were set to obtain an average of 5 mutations per kilobase. 50 bp-long primers were selected to anneal 150 bp upstream and downstream of the REC1-II coding sequence (see Appendix Table 4). The PCR library was directly assembled *in vivo* by co-transformation of the mutagenized amplicon pool with the p415-GalL-Cas9-CYC1t plasmid, previously digested with NcoI and NheI to remove the REC1-II domain, with an insert/plasmid ratio of 3:1. The two 150 bp homology regions at both ends of the amplicons were used by yeast to repair the digested plasmid by homologous recombination, thus incorporating the mutagenized portion. Clones containing mutations in these 150bp flanking regions were probably negatively selected during this *in vivo* assembly step due to loss of complete homology. Nonetheless, these mutations lied outside our region of interest (the REC1-II domain). The mutagenic library was screened concomitantly to its assembly by co-transformation of the fragments in the yACMO-off4 yeast strain stably expressing a sgRNA matching the on-target sequence located in the TRP1 locus. After transformation, the culture was grown overnight in SD medium lacking uracil and leucine (SDlu, for selecting cells carrying both the sgRNA- and Cas9-expressing plasmids) to allow recovery and correct recombination. The next day, Cas9 expression was induced by growing the culture in galactose-containing medium (SRGlu) for 5 hours prior to plating on several selective plates lacking tryptophan and containing low concentrations of adenine (SDluta₅), to discriminate colonies according to the editing status of the TRP1 and ADE2 loci. After 48 hours, TRP1⁺/ADE2⁻ (red) colonies were streaked on selective plate with low adenine and no tryptophan containing galactose and raffinose (SRGluta₅) to keep Cas9 expression constitutively induced and force the generation of off-target cleavages. After further 48 hours of incubation, Cas9-expressing plasmids were extracted from the red-most streaks, corresponding to colonies in which Cas9 cleaved only the on-target site, and the mutations were characterized by Sanger sequencing.

Yeast colony colour analysis and quantification. All plates images were acquired with a Canon EOS 1100D (1/60, f/9.0 and ISO 800) and analysed with OpenCFU³⁵³. For all images an inverted threshold (value = 2) was used with a radius between 8 and 50

pixels. Discrimination between white and red colonies was obtained by computing the average signal in the RGB channels and setting a manual threshold that accurately discriminates between red and white colonies in each experiment.

Mammalian cells and transfections. 293T/17 cells were obtained from the American Type Culture Collection (ATCC) and were cultured in Dulbecco's modified Eagle's medium (DMEM; Life Technologies) supplemented with 10% fetal calf serum (FCS; Life Technologies) and antibiotics (PenStrep, Life Technologies). 293multiGFP cells were generated by stable transfection with pEGFP-IRES-Puromycin and selected with 1 µg/ml of puromycin. 293blastEGFP were obtained by low MOI infection of HEK293T cells with the EGFP-expressing lentiviral vector pAIB-GFP followed by clonal selection with 5 µg/ml of blasticidin. For transfection, 1×10^5 293multiGFP or 293T cells/well were seeded in 24-well plates and transfected the next day using TransIT-LT1 (Mirus Bio) according to manufacturer's protocol with 400-750 ng of Cas9-expressing plasmids and 200-250 ng of sgRNA-expressing plasmids. For transient transfection experiments involving EGFP expression, 100 ng of the pEGFP-N1 plasmid were used. To determine the level of EGFP downregulation by Cas9 after transfection into 293multiGFP, cells were collected 7 days post-transfection and were analysed by flow cytometry using a FACSCanto (BD Biosciences).

Lentiviral vector production and transductions. Lentiviral particles were produced by seeding 4×10^6 293T cells into a 10 cm dish. The day after, the plates were transfected with 10 µg of each lentiCRISPR-based²²³ transfer vector together with 6.5 µg of pCMV-deltaR8.91 packaging vector and 3.5 µg of pMD2.G using the polyethylenimine (PEI) method. After an overnight incubation, the medium was replaced with fresh complete DMEM and 48 hours later the supernatant containing the viral particles was collected, spun down at 500xg for 5 minutes and filtered through a 0.45 µm PES filter. Quantification of the vector titers was performed using the SG-PERT method³⁵⁴. Vectors stocks were conserved at -80°C for future use.

For transductions, 10^5 293blastGFP cells were seeded in a 24-well plate and the next day were transduced with 0,4 Reverse Transcriptase Units (RTU)/well of each vector by centrifuging at 1600xg 16°C for 2 hours. After an overnight incubation, the viral supernatant was removed and the cells were kept in culture for a total of 48 hours

before adding 0,5 µg/ml puromycin selection that was maintained throughout the experiment. To determine the level of EGFP downregulation by Cas9 after infection, 293blastGFP cells were collected at the indicated time-points after transduction and were analysed by flow cytometry using a FACSCanto (BD Biosciences).

Detection of Cas9-induced genomic mutations. Genomic DNA was obtained at 7 days post-transfection, using the QuickExtract DNA extraction solution (Epicentre). PCR reactions to amplify genomic loci were performed using the Phusion High-Fidelity DNA polymerase (Thermo Fisher). Samples were amplified using the oligos listed in Appendix Table 10. Purified PCR products were analysed by sequencing and applying the TIDE tool³¹⁰. To quantify the CCR2-CCR5 chromosomal deletion, a semi-quantitative PCR approach was set-up using primers flanking the CCR5 on-target site and the CCR2 off-target locus (Appendix Table 10). The number of PCR cycles was modulated in order not to reach the amplification plateau. Quantifications were obtained by performing densitometric analyses using the ImageJ software and exploiting the FANCF genomic locus as an internal normalizer.

Western blots. Cells were lysed in NEHN buffer (20 mM HEPES pH 7.5, 300 mM NaCl, 0.5% NP40, NaCl, 1 mM EDTA, 20% glycerol supplemented with 1% of protease inhibitor cocktail (Pierce)). Cell extracts were separated by SDS-PAGE using the PageRuler Plus Protein Standards as the standard molecular mass markers (Thermo Fisher Scientific). After electrophoresis, samples were transferred to 0.22 µm PVDF membranes (GE Healthcare). The membranes were incubated with mouse anti-FLAG (Sigma) for detecting SpCas9 and the different high-fidelity variants, with mouse anti-α-tubulin (Sigma) for a loading control and with the appropriate HRP conjugated goat anti-mouse (KPL) secondary antibodies for ECL detection. Images were acquired using the UVItec Alliance detection system.

Targeted deep-sequencing. Selected off-target sites²⁸⁰ for the VEGFA3 and EMX1 genomic loci, together with their relative on-target, were amplified using the Phusion high-fidelity polymerase (Thermo Scientific) or the EuroTaq polymerase (Euroclone) from 293T genomic DNA extracted 7 days after transfection with wild-type SpCas9 or evoCas9 together with sgRNAs targeting the EMX1 and the VEGFA3 loci, or a pUC empty

vector. Off-target amplicons were pooled in near-equimolar concentrations before purification and indexing. Libraries were indexed by PCR using Nextera indexes (Illumina), quantified with the Qubit dsDNA High Sensitivity Assay kit (Invitrogen), pooled according to the number of targets and sequenced on an Illumina Miseq system using an Illumina Miseq Reagent kit V3 - 150 cycles (150bp single read). The complete primer list used to generate the amplicons is reported in Appendix Table 6.

A reference genome was built using Picard (<http://broadinstitute.github.io/picard>) and samtools³⁵⁵ from DNA sequences of the considered on-/off-target regions. Raw sequencing data (FASTQ files) were mapped against the created reference genome using BWA-MEM³⁵⁶ with standard parameters and resulting alignment files were sorted using samtools. Only reads with mapping quality above or equal to 30 were retained. Presence of indels in each read for each considered region was determined by searching indels of size 1bp directly adjacent to the predicted cleavage site or indels of size ≥ 2 bp overlapping flanking regions of size 5bp around the predicted cleavage site.

GUIDE-seq. We used GUIDE-seq to determine the genome-wide spectrum of the off-target sites relative to the VEGFA site 2 repetitive locus. 2×10^5 293T cells were transfected with 750 ng of a Cas9 expressing plasmid, together with 250 ng of sgRNA-coding plasmid or an empty pUC19 plasmid, 10 pmol of the bait dsODN containing phosphorothioate bonds at both ends (designed according to the original GUIDE-seq protocol³⁰⁵) and 50 ng of a pEGFP-IRES-Puro plasmid, expressing both EGFP and the puromycin resistance gene. The day after transfection cells were detached and selected with 2 μ g/ml of puromycin for 48 hours to eliminate non transfected cells. Cells were then collected and genomic DNA was extracted using the DNeasy Blood and Tissue kit (Qiagen) following the manufacturer's instructions and sheared to an average length of 500bp with the Bioruptor Pico sonication device (Diagenode). Library preparations were performed with the original adapters and primers according to previous work³⁰⁵. Libraries were quantified with the Qubit dsDNA High Sensitivity Assay kit (Invitrogen) and sequenced with the MiSeq sequencing system (Illumina) using an Illumina Miseq Reagent kit V2 - 300 cycles (2x150bp paired-end).

Raw sequencing data (FASTQ files) were analyzed using the GUIDE-seq computational pipeline³⁵⁷. After demultiplexing, putative PCR duplicates were consolidated into single

reads. Consolidated reads were mapped to the human reference genome GrCh37 using BWA-MEM³⁵⁶; reads with mapping quality lower than 50 were filtered out. Upon the identification of the genomic regions integrating double-stranded oligodeoxynucleotide (dsODNs) in aligned data, off-target sites were retained if at most eight mismatches against the target were present and if absent in the background controls. Visualization of aligned off-target sites is available as a color-coded sequence grid.

Results

Design of a reporter yeast strain for the detection of Cas9 activity

We decided to use *Saccharomyces cerevisiae* as an experimental model to develop a directed evolution screen to isolate high-specificity SpCas9 variants. The advantage of using a yeast-based assay platform resides on one side in the similarities that yeast shares with bacteria, such as a fast doubling rate, the possibility to isolate single clones with ease and the availability of fast and reliable transformation protocols; on the other hand, yeast DNA organization and metabolism is similar to the one of higher eukaryotic cells and is often used as a model to study nuclear biology, while this aspect is completely missed in prokaryotes. This is particularly relevant in consideration of the prospective application of the identified variants in human cells, since it makes more likely that the observed behaviour will be conserved also in mammalian systems. We first designed a strategy to generate auxotrophic reporter yeast strains for simultaneously measuring Cas9 on- versus off-target activity by modifying the TRP1 (chromosome IV) and ADE2 (chromosome XV) genomic loci. By using the *delitto perfetto* approach, we substituted the wild-type coding sequences of the two genes with reporter cassettes in which each CDS was split in two halves by the insertion of a specific target sequence that was matched by a guide RNA of interest (on-target) in the case of TRP1 locus, while different sequences containing a single mismatch each in positions spanning the whole length of the target, from more PAM-proximal to more PAM-distal nucleotides, were located in the ADE2 locus (ADE2off1-off4, see Appendix Table 3). A 100 bp duplication was added on both sides of the target sequence and a stop codon was positioned immediately upstream, in between the two homology regions, to ensure premature interruption of translation (**Fig. 16a**). We designed the on- and off-target sequences on the EGFP coding sequence, a widely used and standardized experimental model that allowed us to validate the functionality of the chosen target sites in a reciprocal experiment performed in mammalian cells by using a fluorescence-disruption assay. Co-transfection of 293T cells with pEGFP-N1 together with SpCas9 and the sgRNA-on induced a robust downregulation of EGFP fluorescence. When a sgRNA matching the off1 sequence (sgRNA-off1) was used instead, we still

observed a clear drop in EGFP intracellular levels (**Fig. 16b**). Notably, the mutation characterizing the sgRNA-off1 is located immediately before the PAM (+1 nucleotide), inside the seed sequence, further demonstrating that the general rule of perfect complementarity between the sgRNA and its target in the first 8-12 PAM-proximal nucleotides is not always respected.

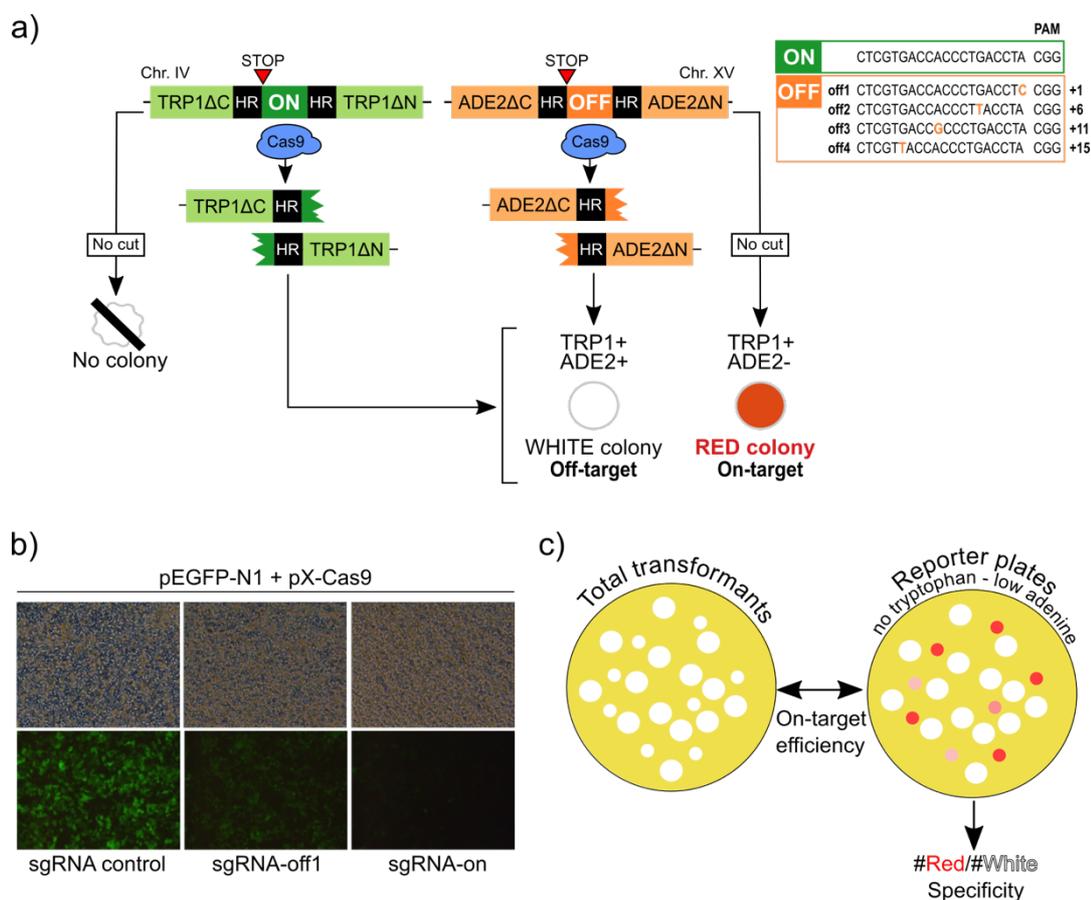


Figure 16. Design of an *in vivo* yeast assay to quantify SpCas9 on- and off-target activity. (a) Generation of yeast reporter strains. The TRP1 and ADE2 loci were modified by the insertion of a reporter cassette containing an on-target site (TRP1) or different off-target sequences (ADE2, sequences are reported in the orange box). The presence of homology regions (HR) on both sides of the target allows efficient repair by single strand annealing upon cleavage by Cas9. Using appropriate selective plates, it is possible to follow the editing status of the two loci. The survival of a colony will indicate TRP1 on-target cleavage, while the colour of the colony allows to assess the cleavage of the ADE2 off-target. (b) SpCas9 has off-target activity. Fluorescence microscope (bottom) and relative bright fields (top) images of 293T cells transfected with EGFP and SpCas9 together with a control sgRNA, a guide perfectly matching EGFP CDS (sgRNA-on) or an off-target sgRNA containing a single mismatch immediately before the PAM sequence (sgRNA-off1). The on and off1 spacers are identical to the ones engineered into the reporter yeast strains. (c) Assay readout. After transformation with SpCas9 and sgRNA coding plasmids cells are plated on selective plates to measure the total number of transformants and on reporter plates with no tryptophan and low adenine to screen for Cas9 editing activity. By comparing the total number of colonies on the two plates the on-target efficiency can be estimated, while by calculating the percentage of white and red colonies on reporter plates a measure of the specificity can be obtained.

We thus generated four different reporter yeast strains that were named yACMO-off1/off4, depending on the off-target sequence characterizing the ADE2 locus. The

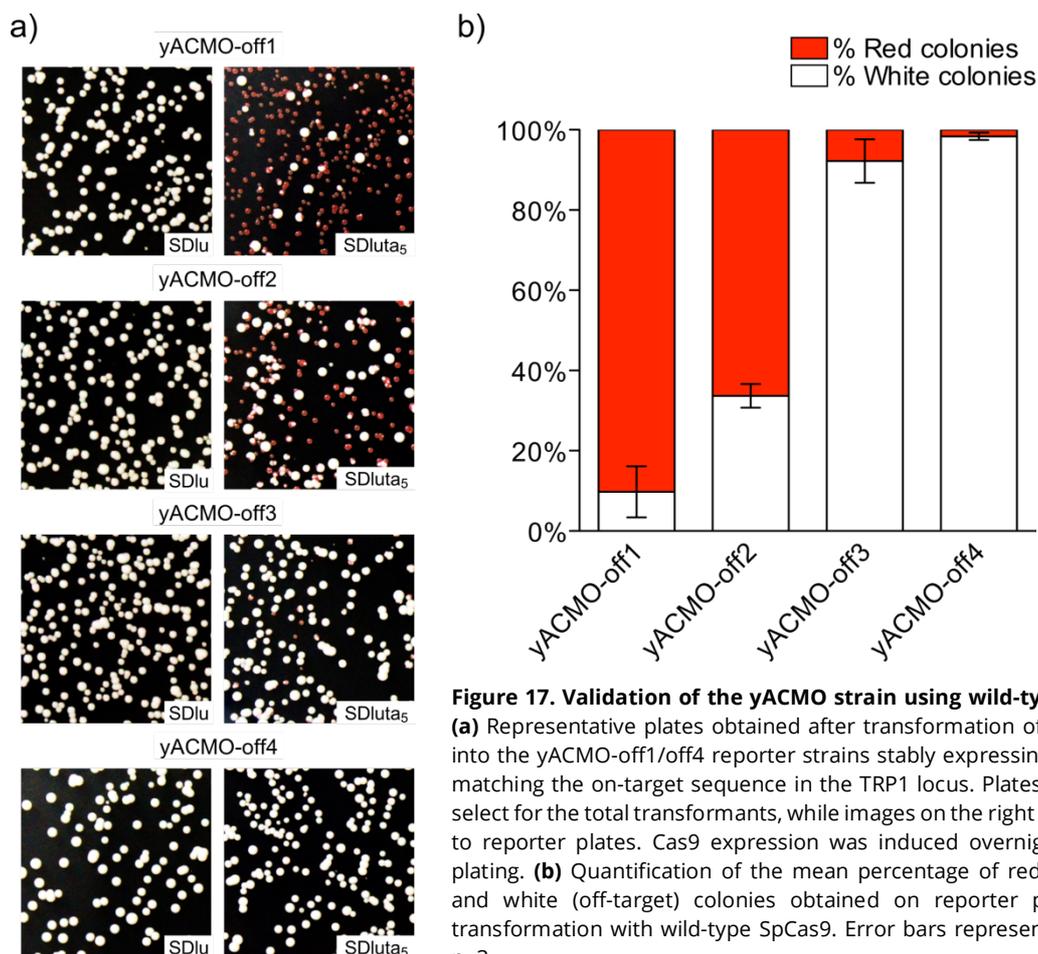
knockout of the TRP1 and ADE2 genes by reporter cassette insertion produces defects in the tryptophan and adenine metabolic pathways, suppressing growth in the absence of tryptophan, and leading to the accumulation in the cell vacuole of an intermediate product of adenine biosynthesis generated by the block at the level of the ADE2 gene product when cells are grown at low concentrations of adenine, thus conferring a characteristic red pigmentation to the colonies on agar plates. Following the formation of double strand breaks induced by Cas9, each locus can be efficiently repaired by yeast cells using the single strand annealing repair pathway thanks to the presence of the two flanking homology regions, obtaining a reversion to prototrophy for the two nutrients. Thus, a screening based on auxotrophies selection can be used to evaluate Cas9 cleavage activity at the two genomic loci, measuring simultaneously both on-targeting and off-targeting events. The successful editing at each of the two loci can be visualized using appropriate reporter plates, which are depleted of tryptophan and contain only low concentrations of adenine (SDluta₅, **Fig. 16c**). After transformation with plasmids for the expression of SpCas9 and the guide RNA, cells are plated in equal numbers on appropriate selective media, to measure the total number of transformants, and on SDluta₅ plates, to discriminate colonies according to the editing status of the off-target locus. The assay readout consists in a two-step process: the first is to compare the number of colonies obtained on SDluta₅ reporter plates and on plates selecting only for the total number of transformants (SDlu plates), lacking uracil and leucine for the auxotrophic selection of the uptake of the sgRNA- and Cas9-expressing plasmids. This allows the evaluation of the on-target cleavage efficiency. The second step consists in counting the number of red colonies (TRP1⁺/ADE2⁻), corresponding to on-target cleavages, and white colonies (TRP1⁺/ADE2⁺), in which also the off-target locus has been edited, on the reporter plates for the evaluation of the on- versus off-target activity (**Fig. 16c**).

Before proceeding, we evaluated the rate of spontaneous reversion for the ADE2 and TRP1 loci of the different yACMO strains that, if too high, could have introduced a confounding variable in the readout of our assay. Reversion of the TRP1 locus can in fact lead to the isolation of false positive clones, while spontaneous recombination of the ADE2 gene can generate false negative colonies. To approximate the experimental conditions used during our assay, we transformed each of the four strains separately

with a plasmid encoding for SpCas9; after 24 hours incubation in selective medium, cells were spread on selective plates to measure the total number of transformants and a 1000-fold more cells were plated on selective plates depleted of tryptophan or adenine, to count the number of revertants for each locus. By comparing the number of colonies obtained in the different selective conditions it was possible to estimate a mean reversion frequency of approximately $1-1,5 \times 10^{-5}$ for both the TRP1 and ADE2 loci.

Validation of the yACMO reporter strain

We validated the functionality of the reporter assay by testing our four reporter strains (yACMO-off1/off4) in combination with wild-type SpCas9. To maximize the overall efficiency, prior to the challenge with SpCas9, each of the strains was stably transformed with a plasmid coding for the sgRNA-on, perfectly matching the on-target sequence in the TRP1 locus. The four strains were then transformed with a plasmid for the expression of wild-type SpCas9 controlled by a galactose-inducible promoter that, after a 4 hour recovery incubation, was induced overnight in galactose-containing



media prior to plating on SDlu and reporter SDluta₅ plates. In these experimental conditions we consistently reached 100% on-target cleavage (compare left and right panels in **Fig. 17a**), while the off-target activity, measured as the percentage of white colonies (TRP1⁺/ADE2⁺) on the reporter plates, increased in accordance with the distance of the mismatched base from the PAM sequence, as expected. For the last two off-targets (off3 and off4), SpCas9 was completely unable to discriminate between the matching and the two mismatched sequences (**Figs. 17a** and **17b**). Considering these results, we decided to screen for SpCas9 variants using the yACMO-off4 strain, containing the strongest off-target sequence, in order to obtain mutants with a marked increase in fidelity.

Yeast-based screening for high-specificity SpCas9 variants

Starting from the assumption that SpCas9 establishes unspecific interactions with the target DNA backbone and that these bonds participate in the stabilization of mismatched gRNA:DNA heteroduplexes, allowing the cleavage of off-target sites, we decided to randomly mutagenize specific Cas9 domains to identify amino acids

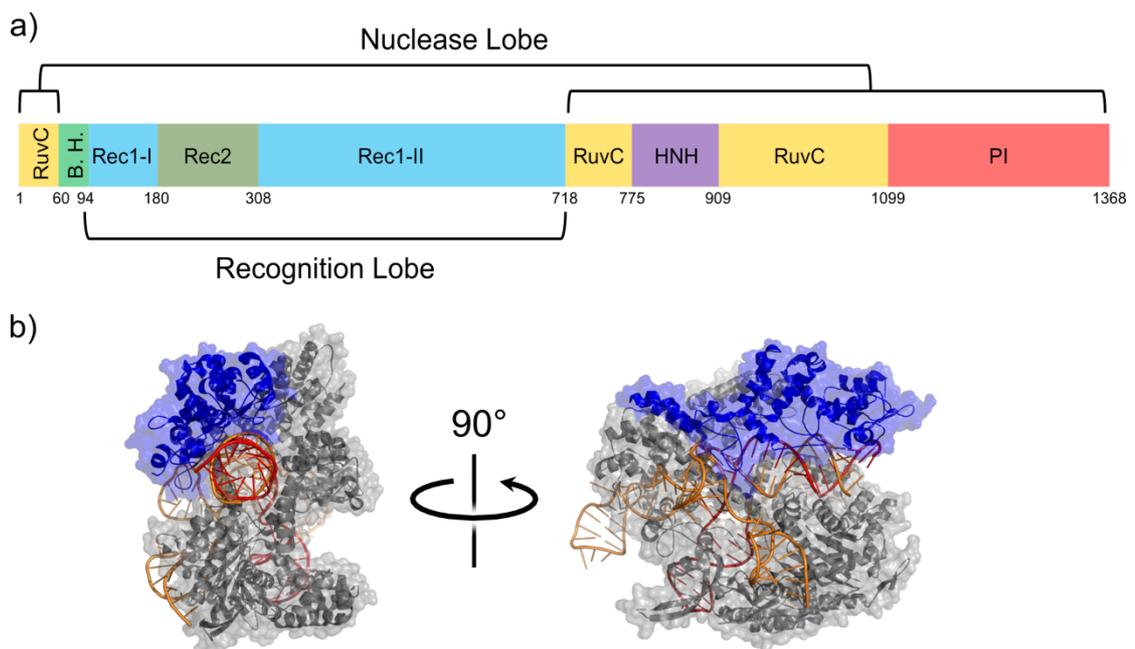


Figure 18. Selection of a target domain for random library generation. (a) Schematics of SpCas9 domains. The Rec1-II domain is part of the alpha helical recognition lobe. BH: bridge helix. **(b)** Hybrid surface/ribbon structure (PDB ID 4UN3) of SpCas9 in complex with a sgRNA and the target DNA. The Rec1-II domain is highlighted in blue, while the rest of the structure is coloured in grey.

substitutions that reduce such unspecific interactions. Differently from published studies^{347,348}, we believed that an unbiased approach could have led to the isolation of non-trivial amino acid substitution increasing the likelihood to obtain a SpCas9 variant with higher fidelity. To find a suitable target for random mutagenesis, we analysed available structural data to identify which domain could be more involved in the formation of such kind of interactions. We excluded from our analysis the nuclease lobe of Cas9, since it contains the two catalytic sites that must be preserved to maintain cleavage activity. The recognition lobe, containing the Rec1, Rec2 and the bridge helix domains, has been reported to make the several contacts with the gRNA:DNA duplex. In addition, the recognition lobe as a whole is one of the least conserved regions across all the three Cas9 families belonging to type II CRISPR systems, indicating a high degree of sequence plasticity. The bridge helix, on the contrary, is one of the most conserved regions among different Cas9 orthologues, suggesting that its sequence is particularly important for nuclease function. The Rec1-Rec2 region spans more than 600 amino acids, a dimension not suitable for random mutagenesis, but the majority of interacting residues are located in the last portion of the Rec1-II domain, approximately between residues 400 and 700 (**Fig. 18**).

A library of REC1-II variants, generated by error-prone PCR to contain approximately 4-5 mutations per molecule, was directly assembled in the yACMO-off4 reporter yeast strain exploiting homologous recombination between the mutagenized REC1-II fragments, containing appropriate homology arms, and a plasmid expressing a galactose-inducible SpCas9 in which the same region had been previously removed. After co-transformation and an overnight recovery incubation in SDlu medium to allow the repair of Cas9-encoding plasmids by homologous recombination and the selection of transformed cells, cultures were induced for 5 hours with galactose to allow SpCas9 expression and cells were plated on several SDluta₅ reporter plates. We shortened the induction time, with respect to previous experiments, to obtain variants that maintained high on-target activity since we observed that wild-type SpCas9 can fully cleave the on-target sequence in this restricted time span (data not shown). Two days later we obtained multiple colonies and streaked the red ones on new reporter plates containing galactose instead of dextrose to reactivate SpCas9 expression and to keep

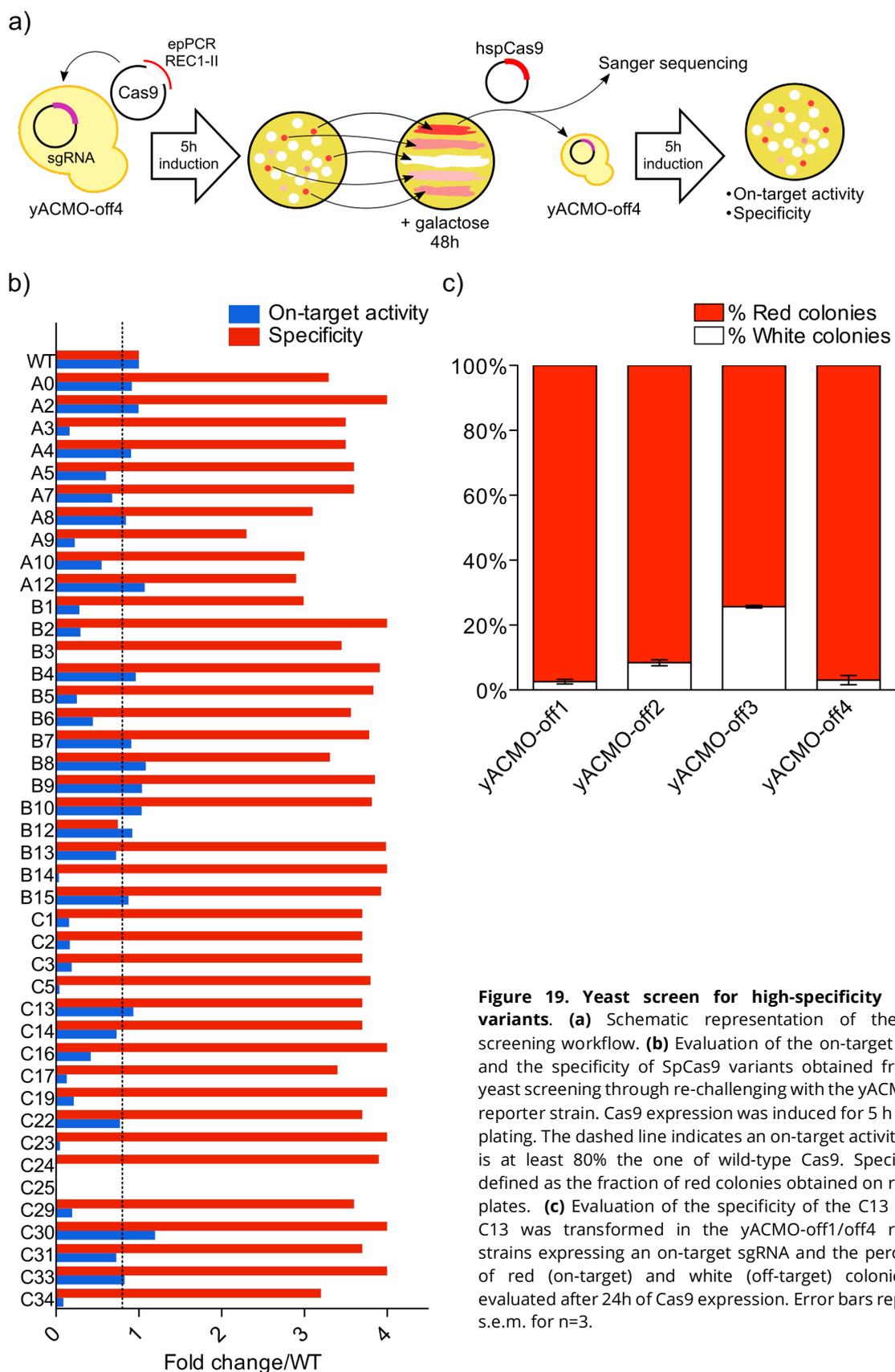


Figure 19. Yeast screen for high-specificity SpCas9 variants. (a) Schematic representation of the yeast screening workflow. (b) Evaluation of the on-target activity and the specificity of SpCas9 variants obtained from the yeast screening through re-challenging with the yACMO-off4 reporter strain. Cas9 expression was induced for 5 h prior to plating. The dashed line indicates an on-target activity which is at least 80% the one of wild-type Cas9. Specificity is defined as the fraction of red colonies obtained on reporter plates. (c) Evaluation of the specificity of the C13 variant. C13 was transformed in the yACMO-off1/off4 reporter strains expressing an on-target sgRNA and the percentage of red (on-target) and white (off-target) colonies was evaluated after 24h of Cas9 expression. Error bars represent s.e.m. for n=3.

it constantly switched on to exacerbate any off-target effect. After 48 hours, plasmids were recovered from the red-most streaks and after amplification in bacteria were Sanger-sequenced to identify the mutations introduced in the REC1-II domain. After a

single round of selection, we identified several amino acidic substitutions, some of which were present more than once in the mutants' pool, in combination with different groups of mutations. Of note, it is likely that mutants containing the same set of variations represent clones deriving from the same original cell that replicated during the recovery incubation. However, given the diversity of substitutions obtained, this phenomenon did not affect the results of our screening. We then performed a re-challenging experiment in the yACMO-off4 strain with each isolated variant in order to measure more precisely its cleavage activity, discard those that did not efficiently cut their target compared to wild-type SpCas9 and rank the remaining ones according to the latter parameter and their ability to discriminate off-target sites (**Fig. 19b**). An outline of the experimental workflow is schematized in **Fig. 19a**. To further validate the results of the screening, we evaluated more in detail the specificity of one of the obtained variants (C13 variant) by challenging all four yACMO reporter strains. After 24 hours of Cas9 expression, the quantification of white and red colonies on reporter plates showed significantly reduced off-target activity when compared to wild-type SpCas9 (compare **Fig. 19c** and **Fig. 17b**).

Optimization of high-fidelity SpCas9 variants in mammalian cells

We selected a pool of substitutions belonging to best performing variants isolated from the yeast screen according both to on-target cleavage efficiency and reduction of unspecific activity. We then reasoned that their hierarchical combination would have allowed to obtain a significant increase in fidelity with respect to the identified mutants, since it could be expected that some of the mutations in each randomly generated variant may have been neutral or detrimental.

Using a reporter cell line stably expressing EGFP (293multiGFP), we tested the on-target activity (sgGFPon) of double mutants (DM) by measuring the loss of fluorescence induced by frameshift mutations into the EGFP coding sequence. In parallel, we evaluated their ability to avoid the cleavage of the same site after the introduction into the sgRNAs of one or two mismatched bases in PAM-distal positions (position 18 for sgGFP18 and positions 18-19 for sgGFP1819). Wild-type SpCas9 was not able to discriminate these two surrogate off-target sequences, as confirmed by an equal cleavage efficiency when guided by both matched and mismatched sgRNAs that

produced the same reduction in the percentage of EGFP⁺ cells (**Figs. 20a** and **20b**). After a first round of selection we combined the top performing substitutions into triple mutants (TM) and repeated the challenging of the EGFP reporter cell line (**Fig. 20a**). We then performed a last round of selection after generating a quadruple mutant by combining the best substitutions of the previous round (Variant A). In addition, we tested another sgRNA containing two mismatches in a more PAM-proximal region (positions 13 and 14, sgGFP1314) to verify that the observed increase in fidelity of our variant was conserved for mutations spanning the whole spacer sequence. Variant A induced little or no loss of EGFP fluorescence for all mismatched guide RNAs, a result that was particularly striking for the sgRNA containing a single substitution in position 18 from the PAM (sgGFP18). On the other hand, this strong increase in specificity produced a small albeit measurable decrease in on-target activity (~20% loss, **Fig. 20b**). In order to address this issue, we generated by rational design two alternative

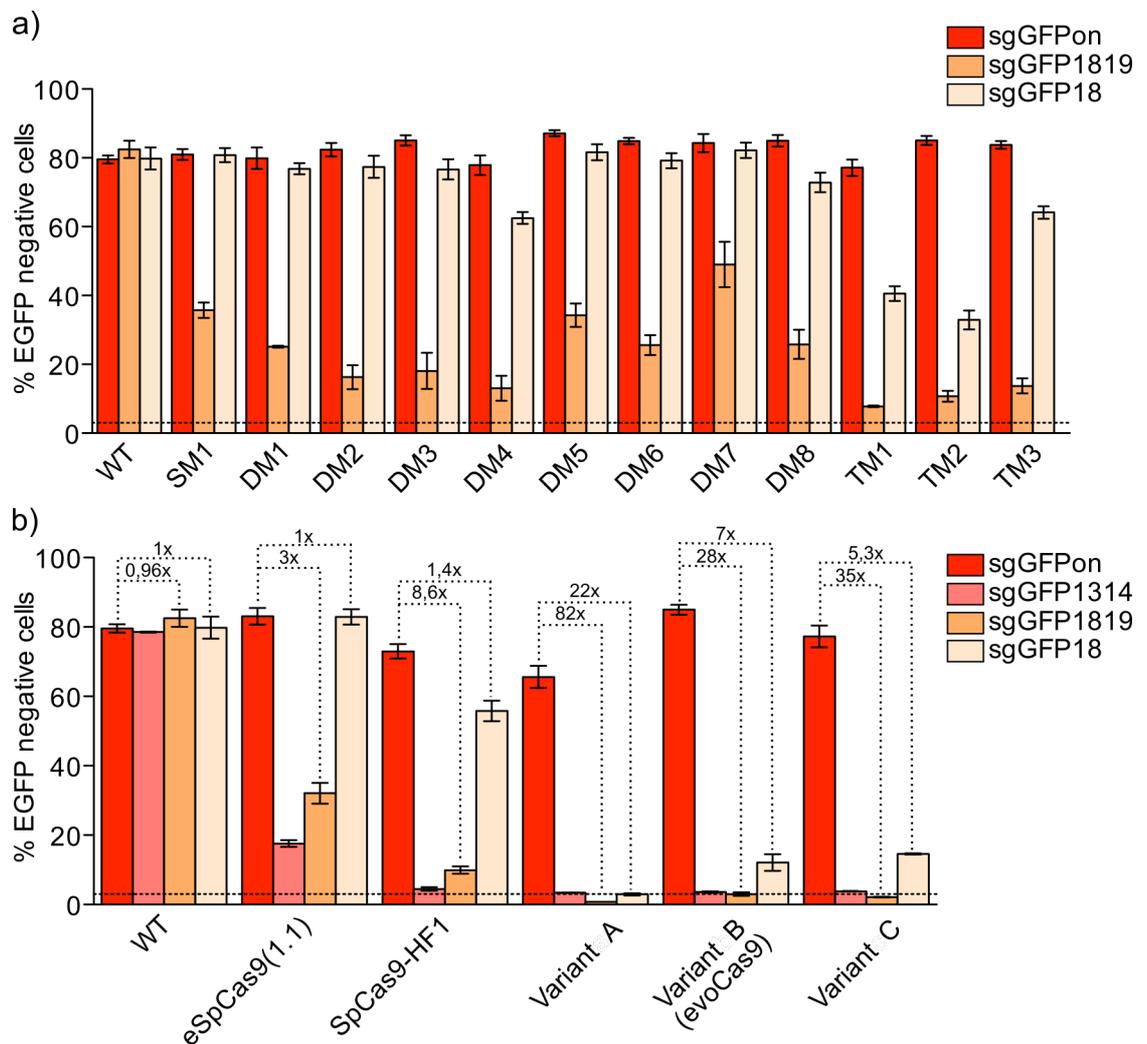


Figure 20 (on previous page). Selection of optimized SpCas9 variants in mammalian cells. (a) Hierarchical combination of mutations obtained from the yeast screening. 293 cells stably expressing EGFP were transfected with single (SM), double (DM) and triple (TM) mutants generated by the hierarchical combination of mutations obtained from the best performing yeast-isolated variants together with an on-target sgRNA (sgGFPon) or each of the mismatched guides. Loss of EGFP fluorescence was measured by FACS analysis at 7 days post-transfection. **(b)** Side-by-side comparison of the best generated variants with previously published mutants. The 293-EGFP knockout assay was used to assess the specificity of the top isolated variants (Variant A-B-C) and to compare their performance with previously published high-fidelity mutants. Loss of EGFP fluorescence was measured by FACS analysis at 7 days post transfection. sgGFP1314 contains mismatches in position 13&14 from the PAM; sgGFP1819 contains mismatches in positions 18&19; sgGFP18 contains a single mismatch in position 18. Dotted lines indicate on/off target ratios calculated for the indicated on/off couples. Dashed lines indicate the background loss of EGFP fluorescence. Error bars represent s.e.m. for $n \geq 2$.

derivatives of Variant A (Variant B and Variant C) and we tested their characteristics using the same EGFP knockout assay. As expected, we observed a complete restoration of on-target cleavage efficiency coupled to a small increase in off-target activity (**Fig. 20b**).

Side-by-side comparison of our quadruple mutants (Variant A-B-C) with previously published high-fidelity variants^{347,348} using our EGFP reporter cell line revealed a marked increase in fidelity, which was particularly evident using the sgRNA containing a single mismatch in position 18. For this particular surrogate off-target, we measured approximately a 17 to 4-fold absolute reduction in unspecific cleavage when comparing Variant A-B-C with SpCas9-HF1, which in our hands was already discriminating mismatched sites much better than eSpCas9(1.1) (**Fig. 20b**). This observation was further confirmed by calculating the on-/off-target ratios of the different SpCas9 variants calculated for the two strongest surrogate off-targets (sgGFP1819 and sgGFP18) (compare dotted lines in **Fig. 20b**).

We next assessed more in detail the on-target activity of Variant A and Variant B by targeting different regions of the EGFP coding sequence (a complete list of the target sites is reported in **Appendix Table 5**) using the 293multiGFP reporter cell line and measuring the loss of fluorescence. We did not analyse Variant C any further since it behaved similarly to Variant B. In accordance with previous results (**Figs. 20a and 20b**), we observed wild-type levels of activity for Variant B, while Variant A was slightly underperforming at some of the sites, with a significant drop in activity for one of the tested loci (**Figs. 21a and 21b**). To rule out the possibility that the different cleavage behaviour measured towards on- and off-target sites was due to an alteration of the intracellular levels of our SpCas9 variants, we analysed by western blot the lysates of 293T cells transfected with equal amounts of wild-type SpCas9, Variant A and Variant

B, together with the two previously published high-fidelity variants eSpCas9(1.1)³⁴⁸ and SpCas9-HF1³⁴⁷. This experiment did not show any major difference in the expression levels of the analysed Cas9 (**Fig. 21c**). Given these results, we decided to further characterize Variant B, which we named evoCas9 (evolved Cas9), that retained near wild-type levels of activity and reduced drastically the cleavage of non-matching sequences in our EGFP-disruption cellular model.

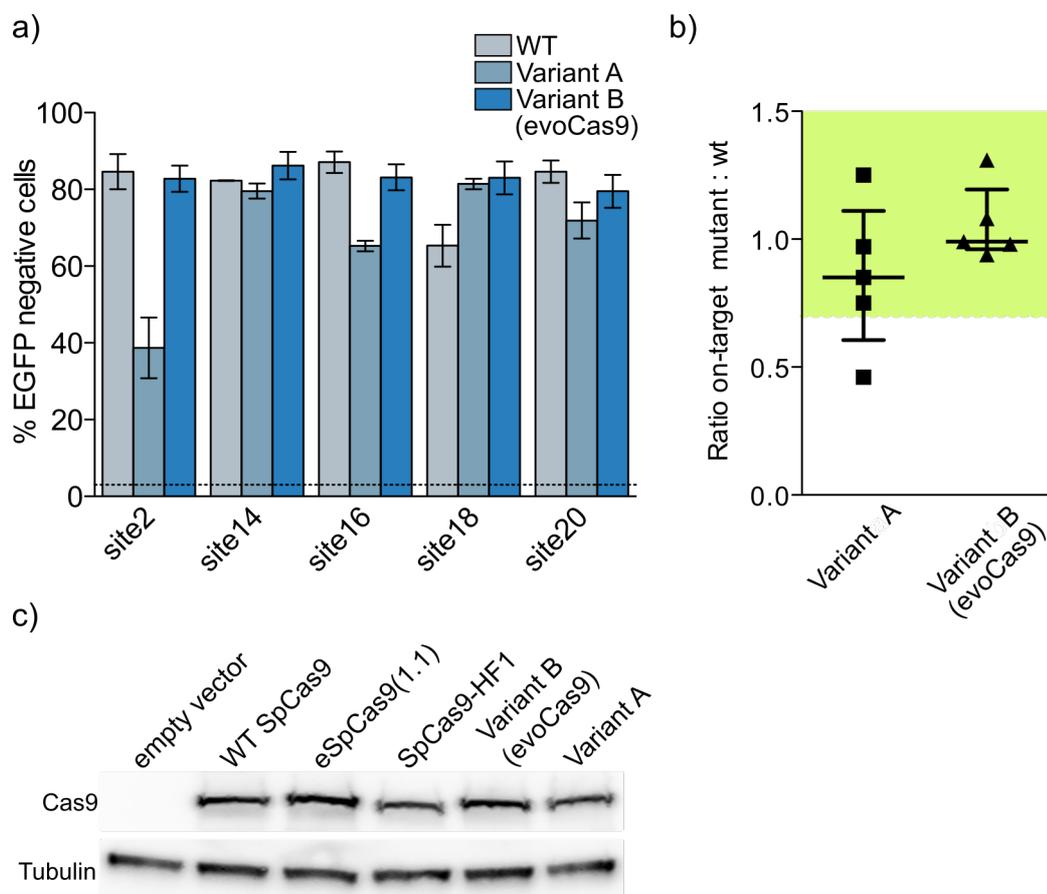


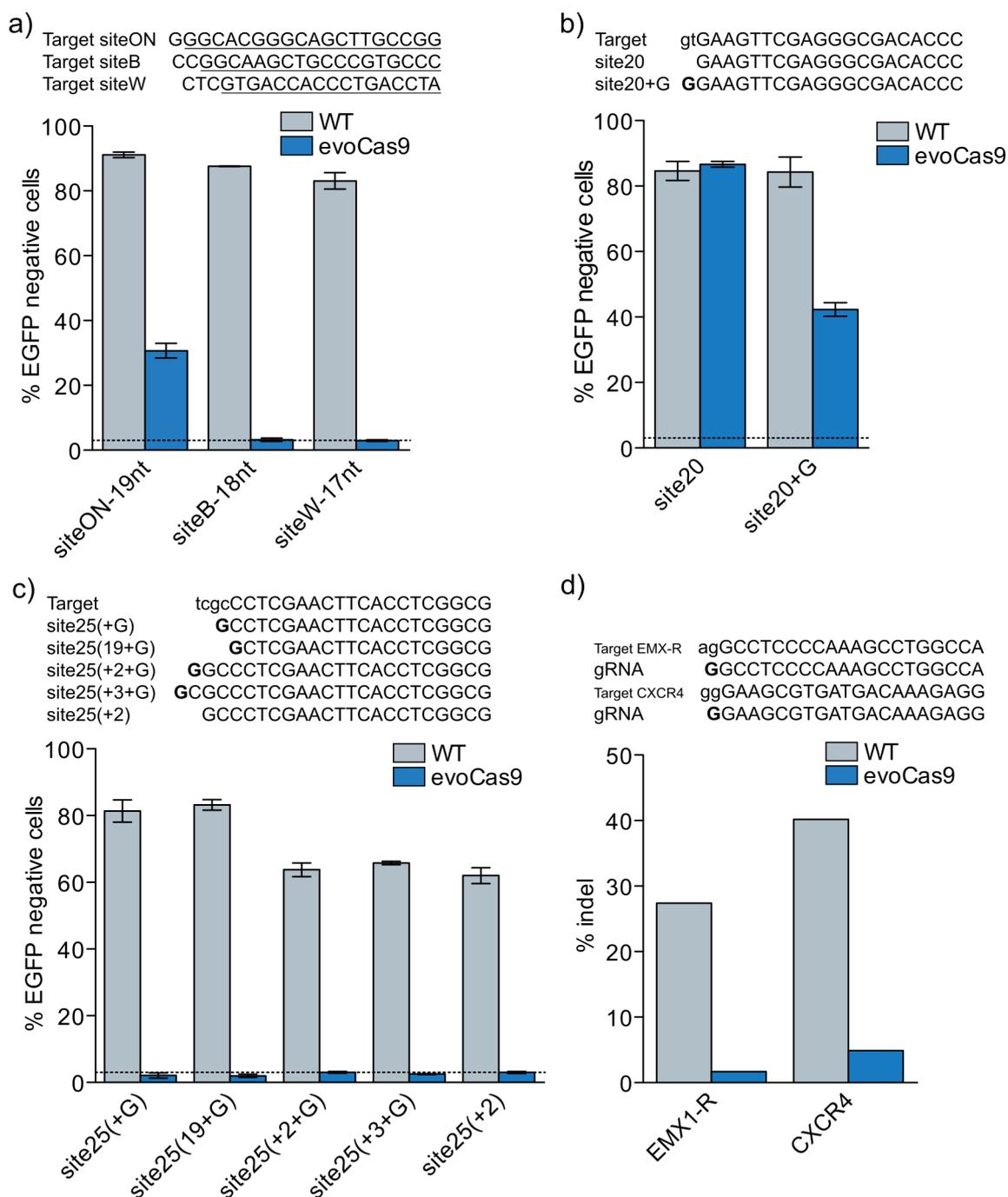
Figure 21. evoCas9 on-target activity against EGFP. (a) evoCas9 activity against EGFP loci. 293 cells stably expressing EGFP were transfected with wt SpCas9, evoCas9 or Variant A together with sgRNA targeting different regions of the EGFP coding sequence. Loss of EGFP fluorescence was measured by FACS analysis at 7 days post transfection. Dashed lines indicate the background loss of EGFP fluorescence. Error bars represent s.e.m. for $n \geq 2$. (b) Ratio of on-target activity of evoCas9 and Variant A to wild-type SpCas9 calculated on EGFP loci. The median and interquartile range are shown. A level of on-target activity above 70% of the wt protein is indicated in green. (c) evoCas9 intracellular expression. Representative western blot of lysates from 293T cells transfected with wt SpCas9, evoCas9 or the other high-fidelity variants. Tubulin was used as a loading control. SpCas9 is detected using an anti-FLAG antibody.

Characterization of sgRNA requirements of evoCas9

Since previous reports³⁴⁷ highlighted the incompatibility of a combination between truncated sgRNAs³²² and high-fidelity SpCas9 variants as a way to further increase their specificity, we decided to test if evoCas9 was able to target different regions of the EGFP

coding sequence using sgRNAs with a length spanning from 17 to 19 nucleotides by co-transfection in the 293multiGFP reporter cell line. As expected, evoCas9 was unable to cleave the two shorter targets and retained only marginal activity with the 19 nt sgRNA, while wild-type SpCas9 was characterized by high activity with all the tested sgRNAs (**Fig. 22a**). This result is in accordance with the increased dependency of our high-specificity SpCas9 variant on the pairing between the sgRNA and the target DNA sequence and suggests that in some cases even a 19 nucleotide pairing between the guide RNA and the target could not be enough to ensure cleavage. We then investigated if the common practice of adding a mismatched G nucleotide at the beginning of sgRNAs to favour transcription from the human U6 polymerase III promoter could be applied to evoCas9 without losing nuclease activity, since previous studies reported such effect³⁴⁷. We thus tested in parallel two sgRNAs targeting the EGFP coding sequence differing only for the presence/absence of a 5'-G and observed that the presence of the additional G in an otherwise matching spacer produced a drop in evoCas9-mediated cleavage, leaving wild-type SpCas9 activity unaffected (**Fig. 22b**). We next generated a set of sgRNAs targeting a single EGFP locus containing 19 to 23 matching nucleotides with an added guanine at the beginning of the spacer sequence and tested their efficacy in our EGFP-knockout model. While wild-type SpCas9 retained robust activity with all the tested guides, even though longer spacers were less efficient in promoting cleavage, evoCas9 was unable to target the EGFP CDS with any of the selected sgRNAs (**Fig. 22c**). This result was further confirmed by measuring the editing of two genomic loci which were targeted using sgRNAs containing an additional G at their 5'-end (**Fig. 22d**). In addition, an sgRNA containing a 22 bp spacer completely matching the target sequence and starting with a matched 5'-G, was able to induce the cleavage of its target only when coupled with wild-type Cas9, but not with evoCas9 (**Fig. 22c**). Notably, the addition of a single mismatch in position 20 of the guide RNA (sgRNA site25(19+G)) completely abrogates evoCas9 cleavage. Taking into account the current working model that assumes that mismatches located in more PAM-distal positions of the spacer sequence are more easily tolerated by Cas9, this result represents an unexpected discrepancy with previous data (see **Fig. 20b**) showing that a mismatch in position 18 of the guide cannot be completely discriminated by evoCas9. This, together with the complete loss of activity observed with spacer sequences longer than 20 bp

and containing or not mismatched 5'-Gs, indicates that evoCas9 does not tolerate alterations of the RNA:DNA heteroduplex in PAM-distal positions that are structurally located at the end of the cavity that harbours the duplex itself. In addition, other groups have reported that the inclusion of two extra guanines at the 5'-terminus of sgRNAs increases targeting specificity²⁹⁸. This effect, which is still without a biological



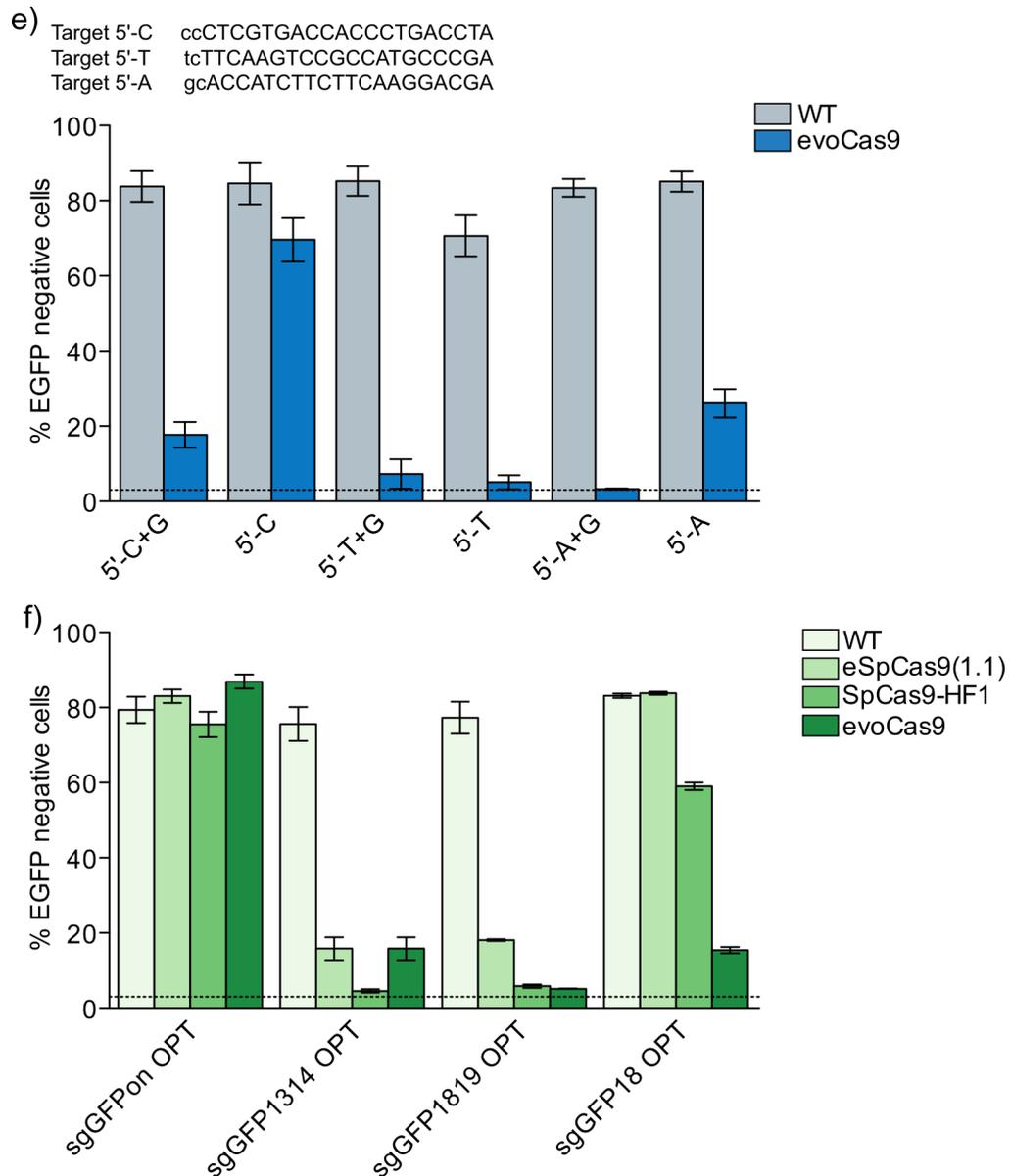


Figure 22 (continued). sgRNA requirements for evoCas9 activity. The EGFP knockout assay was performed using wt SpCas9 and evoCas9 together with different truncated sgRNAs (17-19 nt) (a), an sgRNA containing a mismatched 5'-G nucleotide (site 20+G) or with the same sgRNA without a 5'-G (site 20) (b) and with sgRNAs spanning a length of 19-23 nt and containing a mismatched 5'-G or fully annealing with the target sequence (site 25+2, 22 nt spacer) (c). (d) Indel analysis at two genomic loci targeted using wt SpCas9 and evoCas9 guided by sgRNAs containing a 5' mismatched G. (e) EGFP knockout assay with wt SpCas9 and evoCas9 using sgRNAs with fully matching spacers starting with nucleotides other than G or with their counterparts containing a mismatched 5'-G. (f) evoCas9 activity and specificity using optimized sgRNAs. evoCas9 together with wt SpCas9 and previously published variants were tested using the EGFP disruption assay with optimized versions of the sgRNA presented in Fig. 20 to assess the effect on on-target activity and specificity. Above the graphs further information on the target sequences and the spacers is reported. In (a) underlined nucleotides represent the actual spacer sequence used. Bold Gs indicate non-matching 5' guanines. Lowercase nucleotides represent the bases immediately preceding the target site at each locus. For all the experiments loss of EGFP fluorescence was measured by FACS analysis at 7 days post transfection. For endogenous loci indel analysis cells were collected 7 days post-transfection. Dashed lines indicate the background loss of EGFP fluorescence. Error bars represent s.e.m. for $n \geq 2$.

explanation, in our opinion could be connected to the behaviour we observed with our mutant. Given the impossibility to exploit the common strategy to add an extra G at the 5'-end of the sgRNA when using our high-specificity SpCas9 variant, we next

explored the feasibility of expressing sgRNAs starting with a nucleotide other than G using the human U6 polymerase III promoter. We designed a set of guide RNAs targeting selected EGFP loci with 20 bp spacers starting with C, T and A and a corresponding set in which the same guides were modified by the addition of a 5' mismatched G. When tested in combination with wild-type SpCas9 and evoCas9 using our EGFP-knockout cellular model, we measured robust editing for all the selected guides (both with and without the extra 5'-G) in the presence of wild-type SpCas9, while, on the other hand, evoCas9 was unable to cleave its targets when guided by 5'-G-containing spacers, as expected. Nonetheless, when the 5' extra G was removed, we were able to obtain a partial restoration of the targeting activity for two out of three tested guides (**Fig. 22e**). Overall, these data indicate that it may be possible to evaluate on a case-by-case basis the employment of spacer sequences not starting with a guanine, even without its addition at the beginning of the transcript. Lastly, since the introduction of mutations may have altered SpCas9 contacts with the sgRNA, we tested whether evoCas9, together with eSpCas9(1.1) and SpCas9-HF1, was still able to bind an optimized version of the sgRNA²⁶⁸, which was structurally modified to increase its transcription and interaction with SpCas9 to obtain a more efficient target cleavage. We generated the optimized versions of sgGFPon and its previously employed surrogate off-targets and we tested their behaviour in our EGFP-disruption model. In accordance with the data obtained with non-optimized sgRNAs, we observed a complete absence of discrimination between matched and mismatched spacers by wild-type SpCas9, while the mutated variants could prevent unspecific cleavages, albeit to different extents, similarly to what previously observed (compare **Fig. 22f** with **Fig. 20b**). In particular, this side-by-side comparison confirmed that our high-specificity mutant, evoCas9, was able to better discriminate non-matching spacers, especially when containing a single mismatch in a more PAM-distal position (sgGFP18-OPT, **Fig. 22f**). In addition, we probably did not observe a further increase in on-target cleavage since our reporter assay had already reached a plateau, likely determined by the transfection efficiency relative to the cell line used (**Fig. 22f**). Altogether, these data indicate that structurally modified sgRNAs can be employed in combination with newly generated high-specificity variants without encountering any major negative effect on their enhanced target discrimination capability.

evoCas9 activity towards endogenous loci

We next wanted to further validate our findings using endogenous loci. We thus selected a group of previously tested genomic target sites in order to compare the cleavage activity of evoCas9 with the one of wild-type SpCas9 at each locus. In addition, we introduced in the comparison also SpCas9-HF1, as a further benchmark. After transfection in 293T cells of each SpCas9 variant together with sgRNAs targeting the different loci, we analysed indel formation by using the Tracking of Indels by Decomposition (TIDE) software package on Sanger-sequenced amplicons relative to each target site. For the majority of the loci, we did not observe any major difference in targeting efficiency between wild-type SpCas9 and evoCas9, with the latter being in general slightly less active with an overall mean activity which is 80% of that of the wild-type protein (**Fig. 23a** and **Fig. 23b**). Compared to the other loci, evoCas9 showed very poor cleavage efficiency towards the ZSCAN2 locus (**Fig. 23a**). As regards the SpCas9-

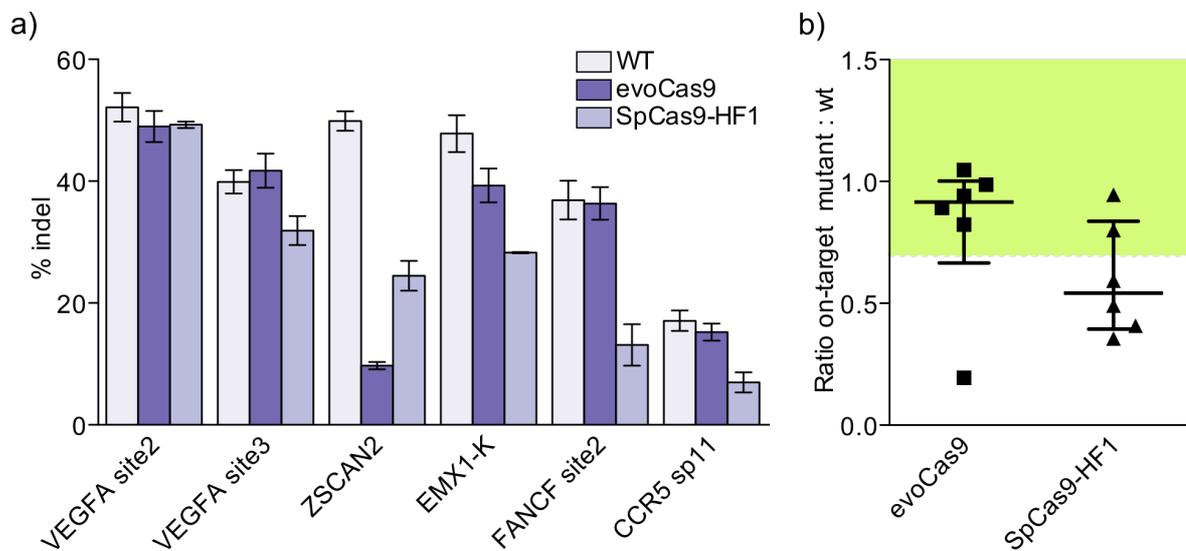


Figure 23. evoCas9 activity on endogenous loci. (a) wt SpCas9, evoCas9 and SpCas9-HF1 activities towards endogenous loci were compared by transfecting 293T cells and by measuring indel formation at 7 days post-transfection using the TIDE tool. (b) Ratio of on-target activity of evoCas9 and SpCas9-HF1 to wild-type SpCas9 calculated on endogenous loci. The median and interquartile range are shown. A level of on-target activity above 70% of the wt protein is indicated in green. Error bars represent s.e.m. for n=2.

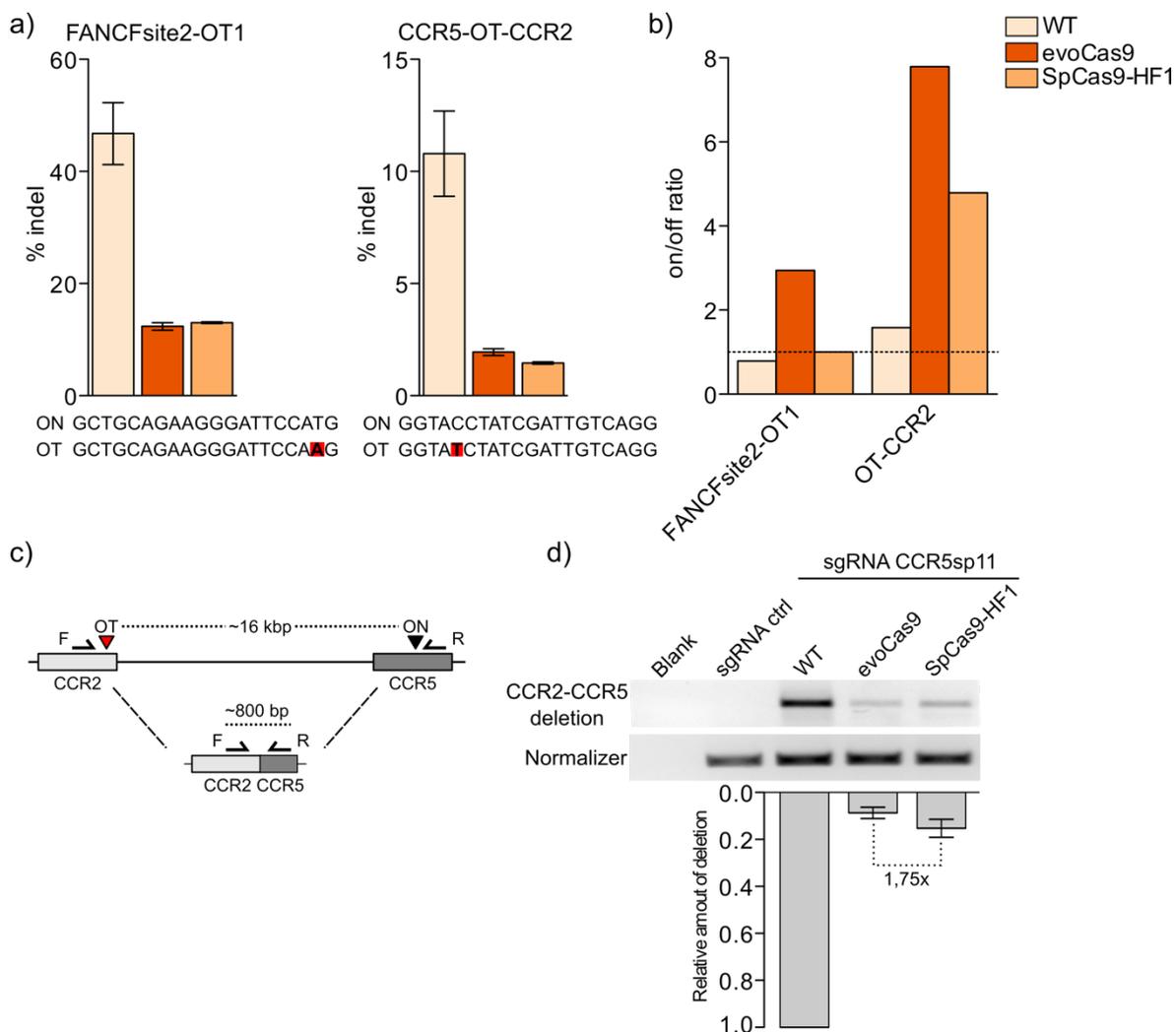
HF1 variant, we measured generally lower cleavage efficiency, with a global mean activity which is 60% if compared to the wild-type (**Fig. 23a** and **Fig. 23b**). This is not in accordance with previously published observations³⁴⁷ and we hypothesize that this discrepancy could be due to the different experimental system used. Nevertheless, since all the experiments have been run in parallel, we believe to have conducted a fair comparison between the two high-fidelity variants. These data demonstrate that

evoCas9 retains near-wild type levels of on-target activity against a panel of endogenous loci, outcompeting the previously reported SpCas9-HF1 variant.

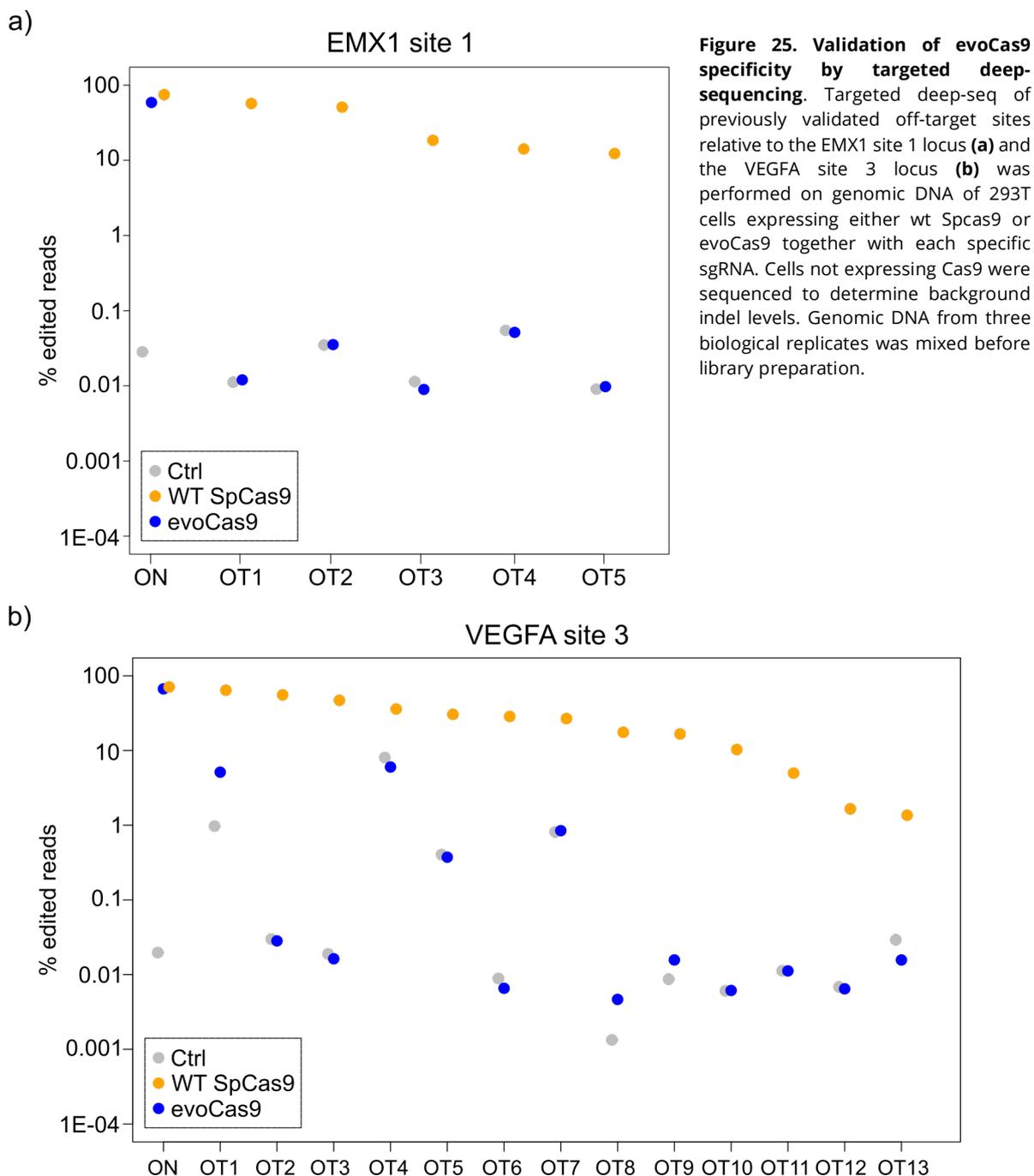
Evaluation of evoCas9 off-target activity

Together with the activity towards on-target sites, we measured evoCas9 specificity by verifying the editing rate at two previously validated off-target sites relative to the FANCF site 2 and CCR5 sp11 loci. In particular, the FANCF site 2-related off-target was one of the few non-repetitive sites that SpCas9-HF1 was unable to discriminate³⁴⁷, while the selected CCR5 sp11 off-target site falls in the highly homologous CCR2 gene, a target that is considered relevant for therapeutic purposes. After transient transfection in 293T cells, we measured indel formation at these two off-target loci using TIDE revealing a significant decrease of cleavage in cells expressing evoCas9 when compared to wild-type SpCas9-transfected cells (**Figs. 24a**). In addition, the calculation of the on-/off-target ratios for wild-type SpCas9, evoCas9 and SpCas9-HF1 confirmed that our variant was able to outperform its competitors at these two loci (**Fig. 24b**). The concerted cleavage of the CCR5 sp11 locus and its off-target in the CCR2 gene can lead to a chromosomal deletion of approximately 15 kilobases (schematized in **Fig. 24c**). We measured the frequency of this chromosomal rearrangement by semi-quantitative PCR in cells transfected with wild-type SpCas9, evoCas9 or SpCas9-HF1 together with the sgRNA targeting the CCR5 locus. While the translocation event was particularly evident in cells edited by wild-type SpCas9, we observed a strong reduction in the amount of deletion to barely detectable levels both in presence of SpCas9-HF1 and evoCas9. Densitometric analyses further revealed that deletion produced by evoCas9 was almost two-fold less with respect to the one generated by the former variant (**Fig. 24d**).

Figure 24 (next page). Side-by-side comparison of evoCas9 and SpCas9-HF1 specificity. (a) Off-target activity of evoCas9 on selected loci. 293T cells were transfected with wt SpCas9, evoCas9 or SpCas9-HF1 together with sgRNAs targeting the FANCF site 2 or the CCR5 sp11 locus. Indel formation at two previously validated off-target sites was evaluated at 7 days post transfection using the TIDE tool. The sequences of the on- and off-target sites for each locus are reported below the corresponding graphs. (b) On/off ratios calculated from the mean indel percentages obtained in (a). (c) Schematic representation of the CCR5 locus and its off-target site in the highly homologous CCR2 gene. Simultaneous cleavage of the two sites generates a chromosomal deletion of approximately 16 kb. (d) Semi-quantitative PCR was performed on genomic DNA of 293T cells transfected with wt SpCas9, evoCas9 or SpCas9-HF1 and the CCR5 guide RNA to assess the amount of chromosomal deletion generated in each condition. The FANCF locus was used as an internal normalizer. The amount of deletion was quantified using densitometry with ImageJ. Error bars represent s.e.m. for n=2.



We then investigated the ability of evoCas9 to avoid unwanted genomic cleavages by performing targeted deep sequencing on a selected set of off-target sites relative to the VEGFA site 3 and EMX1-K genomic loci. All the chosen sites were previously shown to be edited together with the on-target locus²⁸⁰. The advantage of amplicon-seq lies in the possibility to simultaneously measure several targets with high coverage in order to detect even low abundant editing events. Analysis of sequencing data demonstrated that, while retaining high on-target activity on both genomic loci, evoCas9 was able to reduce at background levels the cleavage of the majority of the tested off-target sites (**Figs. 25a and 25b**). The first VEGFA site 3 off-target site (VEGFA3-OT1) was the only locus where we measured editing levels above the background for evoCas9 (**Fig. 25b**). It must be considered, however, that the same locus is edited almost as the on-target site by wild-type SpCas9 and that the previously reported SpCas9-HF1 variant was



unable to completely suppress the unspecific cleavage of this sequence due to its highly repetitive nature³⁴⁷. For four VEGFA site 3 off-target sites (VEGFA3-OT1, -OT4, -OT5, -OT7) we measured significantly higher background editing rates (**Fig. 25b**): we believe that this could be explained by some peculiar characteristics of the local chromatin that is more fragile and thus prone to accumulate mutations.

Altogether these data indicate that evoCas9 significantly decreases unwanted genomic cleavages to undetectable levels for the majority of tested off-target sites. In addition,

side-by-side comparisons with the previously published SpCas9-HF1 variant demonstrated an increased ability to discriminate mismatched sites.

Genome-wide specificity of evoCas9

We decided to extend our analysis of evoCas9 off-target activity at a genome-wide level by using the previously reported GUIDE-seq technique³⁰⁵. This approach is based on the integration of a 34 bp oligonucleotide tag into sites which have been cut by Cas9 in order to allow their capture for library preparation and next-generation sequencing. In this way, it is possible to identify in an unbiased fashion a collection of novel genomic sites associated with a particular guide RNA that are unspecifically targeted by Cas9. We performed a GUIDE-seq analysis of the off-target sites relative to a guide RNA targeting the VEGFA site 2 locus, which is highly repetitive, and has been shown to generate numerous unwanted cleavages into the cellular genome. Additionally, past reports³⁴⁷ indicated that some of the detected off-targets were still cleaved by the high-fidelity SpCas9-HF1 variant. We thus generated GUIDE-seq libraries from genomic DNA of 293T cells transfected either with wild-type SpCas9 or evoCas9, together with the VEGFA site 2 sgRNA and the bait double stranded oligonucleotide. Sequencing data were analysed using a publicly available software pipeline³⁵⁷ revealing for wild-type SpCas9 600 different off-target sites characterized by 7 or less mismatches with the on-target sequence (**Figs. 26a** and **26b**). Of note, approximately 100 of these off-target sequences were edited more efficiently than the on-target site, according to the number of reads obtained by GUIDE-seq analysis that have been reported to be a good proxy of the actual cleavage activity at each specific site³⁰⁵ (**Fig. 26a**). When the same analysis was performed on evoCas9 samples, a total of only 10 sites were detected, the majority of which shared high similarity with the on-target sequence and were characterized by less than five mismatches with the VEGFA site 2 locus (**Fig. 26a**). We detected only one off-target that was cleaved efficiently (more than the on-target) by evoCas9 and this was probably due to the particular nature of this sequence. In fact, it differed by only two nucleotides from the intended target and contained two uninterrupted stretches of cytosines at the level of each mismatch (**Fig. 26a**), maybe allowing the formation of bulge sites to accommodate the non-matching nucleotides.

Overall our GUIDE-seq analysis demonstrated that evoCas9 retains very high-specificity at the genome-wide level even when tested using a repetitive target sequence characterized by multiple off-targets into the cellular genome.

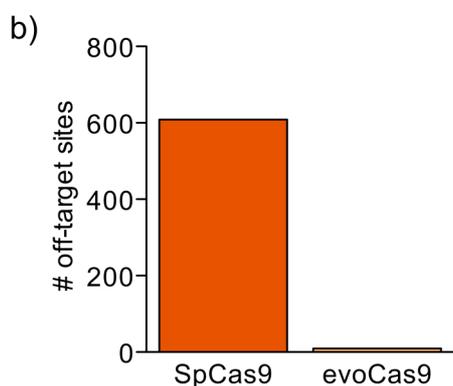
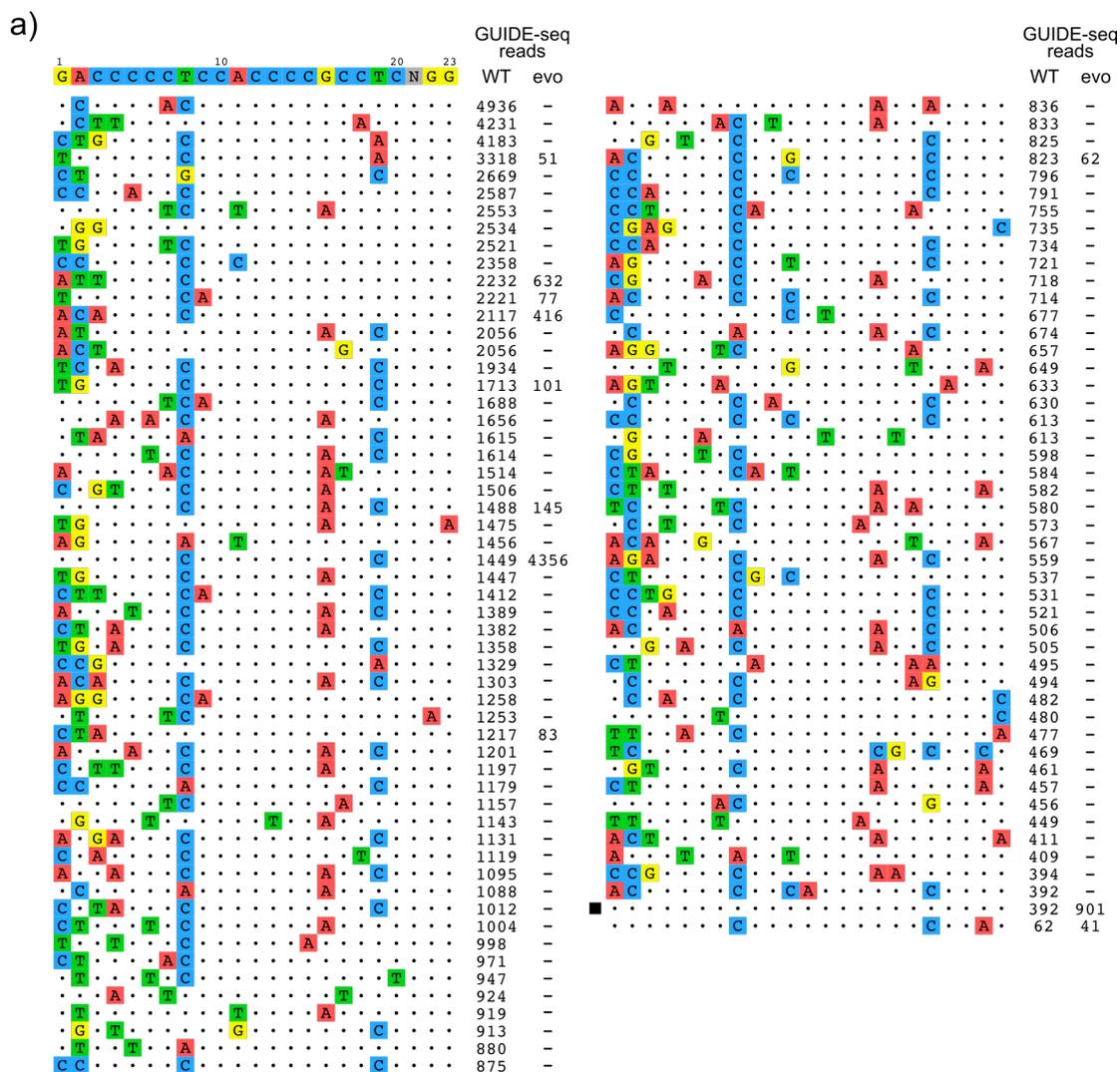


Figure 26. Genome-wide specificity of evoCas9. (a) GUIDE-seq analysis of the off-target sites relative to the VEGFA site 2 locus performed for both wt SpCas9 and evoCas9 in 293T cells. **(b)** Total number of detected off-target sites for wt SpCas9 and evoCas9. Genomic DNA from three biological replicates was mixed before library preparation. Black square denotes on-target site.

Specificity of an evo-dCas9-based transcriptional activator

An alternative application for Cas9, independent from its nuclease activity, is the generation of RNA-guided transcriptional activators by fusing a catalytically inactive

version of Cas9 (dCas9) to various protein domains that stimulate transcription. We built a VP64-based transcriptional activator using a catalytically inactive mutant of evoCas9 (evo-dCas9) and tested it side-by-side with a wild-type dCas9-VP64 activator using a reporter system based on an inducible EGFP expression vector normally regulated by a Tet trans-activator through its binding to an array of Tet operators in the context of a minimal CMV promoter. We substituted the Tet trans-activator with the Cas9-based transcriptional activator guided by an RNA targeting the repeated Tet operator sequences (**Fig. 27a**). We designed two different on-target sgRNAs differing only for the presence or the absence of an extra 5'-G nucleotide, plus three additional mismatched guides based on the same on-target sequence, bearing either one or two mutations in different positions along the spacer sequence, without an added guanine at the beginning of the transcript. We observed lower absolute levels of EGFP fluorescence when the activation was driven by evo-dCas9-VP64 if compared to dCas9-VP64 for both matching and mismatched guide RNAs, suggesting stronger binding to the target DNA by wild-type dCas9 (**Fig. 27b**). However, the on-target fold activation relative to the control sgRNA was similar for both wild-type dCas9 and evo-dCas9, due to the lower background activation observed in the presence of our high-specificity variant, most likely due to the lower propensity of evo-Cas9 to bind DNA unspecifically (**Fig. 27c**). Interestingly, we did not observe any difference in EGFP transcriptional activation when comparing samples transfected with evo-dCas9-VP64 together with the on-target TetO-on guide RNA or the on-target guide with the added extra 5'-G (**Fig. 27b**). This, combined with previous data showing a consistent loss of editing activity when using sgRNA containing an additional starting G (**Fig. 22a-d**), suggests that evoCas9 is indeed able to bind target sites using sgRNAs characterized by a mismatched extra 5'-guanine, but is then unable to cut the bound DNA. Accordingly, when we compared the EGFP fold-activation obtained using mismatched guides, the increased specificity observed using evo-dCas9-VP64 was modest when compared with dCas9-VP64, further reinforcing the idea that evoCas9 binds to mismatched targets, even though less efficiently, but is then unable to complete the cleavage reaction (**Fig. 27c**). Overall, these results indicate that evoCas9 can be exploited to build a transcriptional regulator characterized by lower background activation, albeit less absolute potency, and slightly increased fidelity.

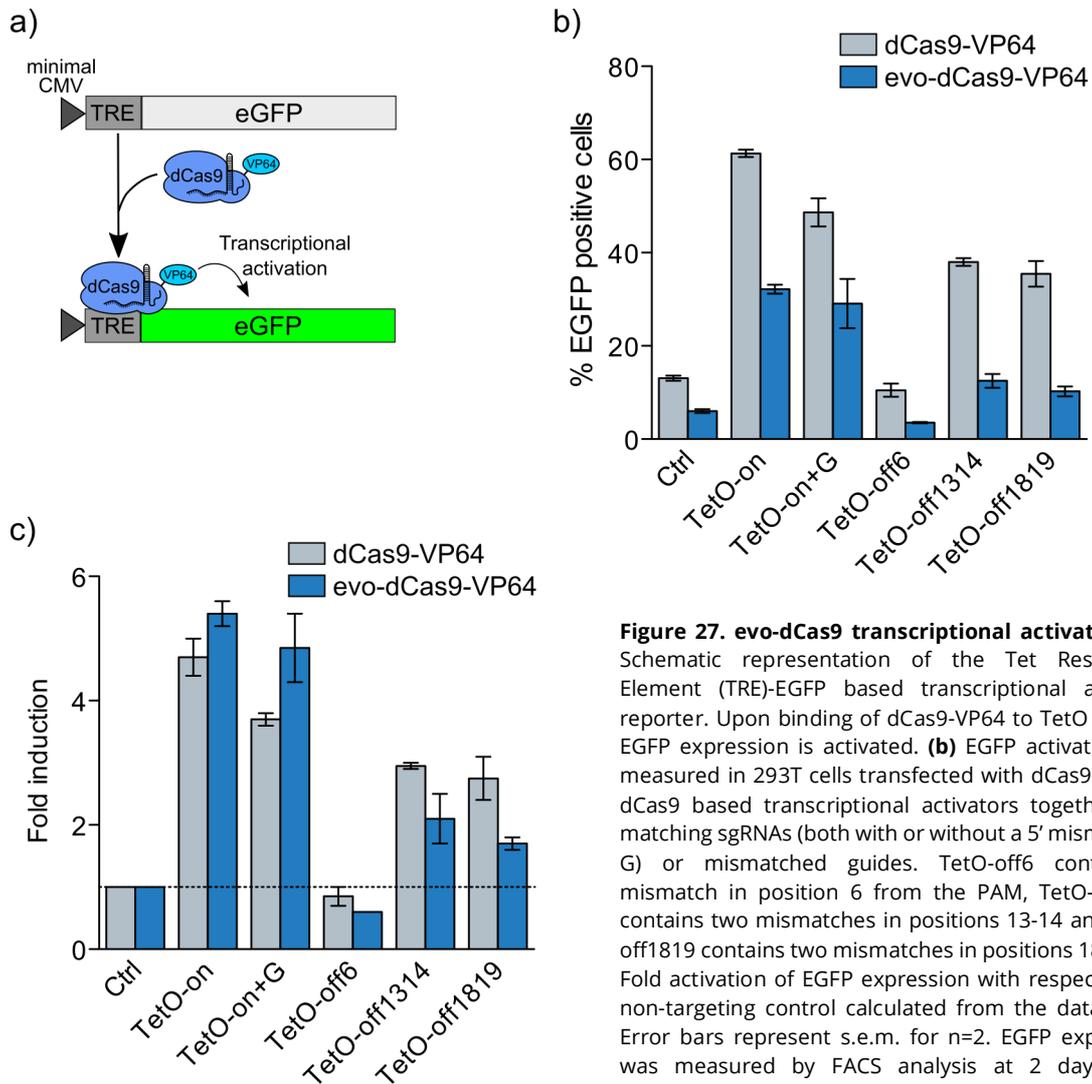


Figure 27. evo-dCas9 transcriptional activation. (a) Schematic representation of the Tet Responsive Element (TRE)-EGFP based transcriptional activator reporter. Upon binding of dCas9-VP64 to TetO repeats EGFP expression is activated. **(b)** EGFP activation was measured in 293T cells transfected with dCas9 or evo-dCas9 based transcriptional activators together with matching sgRNAs (both with or without a 5' mismatched G) or mismatched guides. TetO-off6 contains a mismatch in position 6 from the PAM, TetO-off1314 contains two mismatches in positions 13-14 and TetO-off1819 contains two mismatches in positions 18-19. **(c)** Fold activation of EGFP expression with respect to the non-targeting control calculated from the data in (b). Error bars represent s.e.m. for n=2. EGFP expression was measured by FACS analysis at 2 days post-transfection.

Long-term specificity of evoCas9

Since the permanent expression of Cas9 into cells has been associated with increased off-target activity, an important issue that had to be addressed was the long-term behaviour of evoCas9 into cells. We therefore generated lentiviral particles to obtain the stable expression of wild-type SpCas9, evoCas9 or SpCas9-HF1 together with an sgRNA of interest. We decided to exploit a cellular EGFP-knockout model similar to the one previously employed in this study, together with the same set of sgRNAs directed towards the EGFP coding sequence, either perfectly matching the target locus or containing one or more mismatches in different positions of the spacer sequence, thus acting as surrogate off-targets. We transduced the reporter cell line with equal amounts of the different lentiviral vectors and kept the cultures under antibiotic

selection throughout the duration of the experiments to isolate the transduced population and avoid the eventual dilution of editing events in time due to loss of edited cells or reduced fitness of transduced cells. The measurement of the decrease of EGFP fluorescence at different time points revealed that wild-type SpCas9 cut the target sequence with very high efficacy regardless whether the sgRNA was perfectly matching or not the target site; this was observed even from the earliest time points, in accordance with the results obtained in transient transfection experiments (see **Fig. 28a** and **Fig. 20a-b**). On the contrary, both evoCas9 and SpCas9-HF1 did not cleave EGFP efficiently using the sgGFP1314 and sgGFP1819 sgRNAs, both containing double mismatches (**Fig. 28b-c**). Of note, while loss of fluorescence was at background levels for evoCas9 samples at all time points, a measurable number of EGFP-negative cells that remained constant over time was present in the cultures transduced with SpCas9-HF1-expressing vectors (**Fig. 28c**). In addition, the two Cas9 variants significantly differed when tested with the strongest surrogate off-target, sgGFP18, containing a single mismatch in a PAM-distal position. While SpCas9-HF1 cleaved the locus with similar efficiency with both the matching and mismatched sgRNA, evoCas9 showed a differential activity with the two tested guides. The curves representing the percentage of non-fluorescent cells reported for evoCas9-transduced cultures revealed that the loss of fluorescent cells increased in time and reached less than half the level of the on-target knockout at 40 days post-transduction (**Figs. 28b-c**). Interestingly, we observed a progressive reduction in the number of EGFP negative cells in cultures treated with SpCas9-HF1 (see samples sgGFPon and sgGFP18 in **Fig. 28**). This effect could be connected to Cas9-induced cellular toxicity with the consequential death of highly transduced cells, which would more likely belong to the EGFP-negative population. These data suggest that, even though it is conceivable that not all off-target sites could be avoided by our high-specificity Cas9 variant, a carefully selected guide RNAs combined with evoCas9 may allow long-term expression of the nuclease without any unwanted cleavage occurring in the cellular genome.

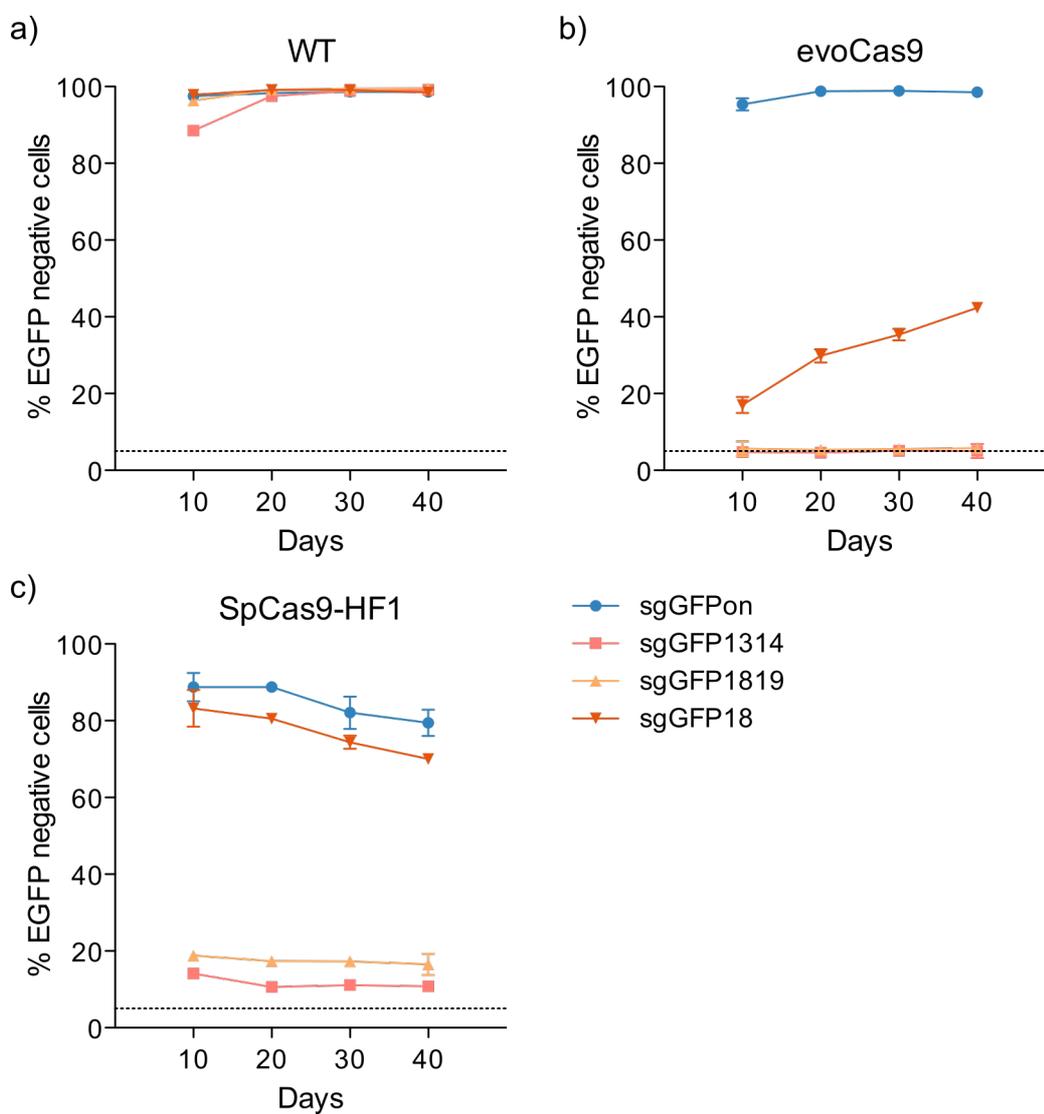


Figure 28. Long-term specificity of evoCas9. Stable expression of wt SpCas9 **(a)**, evoCas9 **(b)** and SpCas9-HF1 **(c)** was obtained through lentiviral transduction of a 293T cell line stably expressing EGFP. Each lentiviral vector was also expressing an on-target sgRNA or the different mismatched guides presented in Fig. 20. EGFP knockout was evaluated by FACS analysis at the time points indicated in the graphs. Cells were kept under puromycin selection for the whole duration of the experiments. Error bars represent s.e.m. for n=2.

Hit-and-go Cas9 delivered through a lentiviral based self-limiting circuit

Rationale of the experiments

An important factor influencing the number of off-target modifications induced by Cas9 is the amount and persistence of its expression in target cells: high concentrations of the nuclease have been reported to increase off-site cleavage, whereas lowering the amounts of SpCas9 increases specificity^{300-302,342}. Transient SpCas9 expression is indeed sufficient to permanently modify the target genomic locus with decreased off-target activity as demonstrated by the enhanced specificity obtained through direct delivery of recombinant SpCas9-sgRNA complexes into target cells³³²⁻³³⁴ or by using a SpCas9 variant activated by inteins, whose expression can be modulated using a small molecule ligand³⁴². It is likely that any Cas9 protein present after the target locus has been edited has a substantial probability to modify additional unwanted sites. Even though direct delivery of SpCas9-sgRNA ribonucleoprotein complexes may decrease off-target effects, it is highly inefficient and unsuitable for *in vivo* approaches. On the other hand, although viral vectors are optimal delivery tools, they generate stable expression of the transferred factors which is not necessarily beneficial for CRISPR/Cas9 applications. We thus combined the strong points of these two technologies by generating a Self-Limiting Cas9 for Enhanced Safety (SLiCES) deliverable through a lentiviral vector. This circuitry, once deployed into target cells, generates a transient Cas9 expression-peak that allows modification of the intended genomic locus, while simultaneously neutralizing Cas9 expression itself, thereby protecting the host cell from long-term nuclease-mediated off-target editing.

Methods

Plasmids and oligonucleotides. The 3XFLAG-tagged *S. pyogenes* Cas9 was expressed from the pX-Cas9 plasmid, which was obtained by removal of an NdeI fragment including the sgRNA expression cassette from pX330-U6-Chimeric_BB-CBh-hSpCas9 (Addgene #42230)²²³. The sgRNAs were transcribed from a U6 promoter driven cassette, derived from pX330 and cloned into pUC19. sgRNA oligos were cloned using a double BbsI site inserted before the sgRNA constant portion following a previously published cloning strategy²²³. For the preparation of optimized sgRNAs the same design strategy was adopted, swapping the constant portion of the guide RNA with its optimized counterpart²⁶⁸. The human codon optimized coding sequence of *S. thermophilus* CRISPR 1 Cas9 used for generating a plasmids expressing a FLAG-tagged St1Cas9 was obtained from pMSP1673²⁸⁰ (Addgene #65769) through Addgene, together with a plasmid to express the *S. thermophilus* gRNA (pBPK2301²⁸⁰, Addgene #65778). *S. thermophilus* spacer oligos were cloned into pBPK2301 using BsmBI and sgRNAs were transcribed from a U6 promoter. The list of sgRNAs target sites employed in this study is available in the Appendix. pcDNA5-FRT-TO-EGFP plasmid was obtained by subcloning EGFP from pEGFP-N1 into pcDNA5-FRT-TO (Invitrogen). pcDNA5-FRT-TO-EGFP-Y66S was obtained by site directed mutagenesis of pcDNA5-FRT-TO-EGFP to insert the Y66S mutation in the EGFP coding sequence. A sgRNA resistant, non-fluorescent truncated EGFP fragment (1-T203K-stop), obtained by site directed mutagenesis of the pcDNA5-FRT-TO-EGFP plasmid, was amplified by PCR and inserted in place of the EGFP in the pcDNA5-FRT-TO-EGFP plasmid, yielding the donor plasmid pcDNA5-FRT-TO-rEGFP(1-T203K-stop).

The SV5-EGFP-based NHEJ reporters employed in this study (Rep. SV5, NHEJ-REP.W and NHEJ-Rep.M) were generated by cloning into the NheI/BspEI sites of the reporter plasmid dsDNA oligos corresponding to the complete target sequence (including PAM) recognized by a sgRNA of interest. The target is inserted between the SV5 tag and EGFP coding sequences, with the EGFP sequence positioned out of frame with respect to the starting ATG codon of the SV5 tag open reading frame (ORF). A stop codon is inserted in the SV5 frame, immediately after the target sequence. The target sequences cloned

in the reporter are listed in Appendix Table 8-9. The pcDNA3 MHC-I-roTag plasmid is described in³⁵⁸.

All the spacer sequences employed in the study are listed in Appendix Table 8 and Table 10.

Cell culture and transfections. 293T/17 cells were obtained from ATCC. 293TR cells, constitutively expressing the Tet repressor (TetR), were generated by lentiviral transduction of parental 293T/17 cells using the pLenti-CMV-TetR-Blast vector (Addgene # 17492)³⁵⁹ and were pool selected with 5 µg/ml of blasticidin (Life Technologies). Similarly, 293T-SpCas9 cells, stably expressing SpCas9, were obtained by lentiviral transduction with the lentiCas9-Blast vector (a gift from Feng Zhang, Addgene # 52962)³⁶⁰ at MOI of 1 and were pool-selected using 5 µg/ml of blasticidin. 293multiEGFP cells were generated by stable transfection of pEGFP-IRES-Puromycin and selected with 1 µg/ml of puromycin. 293-iEGFP and 293-iY66S cells (Flp-In T-Rex system; Life Technologies) were generated by Flp-mediated recombination using the pcDNA5-FRT-TO-EGFP or the pcDNA5-FRT-TO-EGFP-Y66S as donor plasmids, respectively, in cells carrying a single genomic FRT site and stably expressing the tetracycline repressor (293 T-Rex Flp-In, cultured in selective medium containing 15 µg/ml blasticidin and 100 µg/ml zeocin, Life Technologies). Briefly, the parental commercial cell line (293 T-Rex Flp-In) was co-transfected with each of the two donor plasmids together with a plasmid coding for the Flp recombinase (pOG44) in a ratio 1:9 to allow site specific integration of the desired sequence. Antibiotic selection was relaxed for two days to promote recombination before switching to a new antibiotic cocktail (15 µg/ml blasticidin and 100 µg/ml hygromycin, Life Technologies) that selects the correct integration event and allowed the generation of the clonal lines 293-iEGFP and 293-iY66S, which were cultured in the same selective medium. Specific integration in 293-iEGFP and 293-iY66S clones was verified by checking for the loss of zeocin resistance. All cell lines were cultured in DMEM supplemented with 10% FBS, 2mM L-Gln, 10 U/ml penicillin, and 10 µg/ml streptomycin and the appropriate antibiotics indicated above.

293T, 293-iEGFP or 293-iY66S cells were transfected in 12 or 24 multi wells with 250-500 ng of Cas9-expressing plasmid and 250-500 ng of the desired sgRNA-expressing

plasmid using TransIT-LT1 (Mirus Bio), according to manufacturer's instructions. Cells were collected 2-4 days after transfection or as indicated.

In 293-iEGFP and 293-iY66S cells the expression of EGFP was induced by treatment with 100 ng/ml doxycycline (Cayman Chemical) for 20 h before fluorescence measurement.

lentiSLiCES vectors. lentiSLiCES was prepared from the lentiCRISPRv1²²³ transfer vector by substituting the EFS-SpCas9-2A-Puro cassette with a SpCas9(intron)-IRES-Blasticidin fragment together with a CMV-TetO promoter. The intron introduced in SpCas9 derives from the mouse immunoglobulin heavy chain precursor V-region intron (GenBank ID: M12880.1) and is necessary to avoid leaky Cas9 expression during plasmid amplification in bacteria that would lead to plasmid digestion due to the concomitant presence of the self-targeting sgRNA. An EMCV-IRES regulating the translation of a blasticidin resistance gene was cloned downstream of SpCas9 to allow the antibiotic selection of transduced cells, even after the generation of frameshift mutations in the Cas9 coding sequence following Cas9 self-cleavage of the integrated vector. The sgCtr-opt or the sgCas9-a-opt were assembled with an H1-TetO promoter within a pUC19 plasmid, PCR amplified and then cloned into a unique EcoRI site in lentiCRISPRv1 and selected for the desired head-to-tail orientation. The spacer sequences targeting the chosen locus were cloned into the lentiCRISPRv1 sgRNA cassette using the two BsmBI sites, following standard procedures²⁴⁵. All the primers employed for cloning are listed in Appendix Table 7.

Lentiviral vector production. Lentiviral particles were obtained by using 293T or 293TR cells, for lentiCRISPR or lentiSLiCES production, respectively. Lentiviral vector production and collection has been described in the Methods section of the first part of the results. After collection, lentiSLiCES viral vectors were concentrated using polyethylene glycol (PEG) 6000 (Sigma). Briefly, a 40% PEG 6000 solution in water was mixed in a 1:3 ratio with the vector-containing supernatant and incubated for 3 hours to overnight at 4°C. Subsequently, the mix was spun down for 45 minutes at 2000xg in a refrigerated centrifuge. The pellets were then resuspended in a suitable volume of DMEM complete medium. lentiCRISPR vectors were used unconcentrated. The titer of

the lentiviral vectors (reverse transcriptase units, RTU) was measured using the product enhanced reverse transcriptase (PERT) assay³⁵⁴.

Infections and EGFP fluorescence detection. One day before transduction 10^5 293T, 293-iEGFP or 293-multiEGFP cells were seeded in a 24-well plate. For lentiSLiCES vectors, cells were transduced by centrifuging 2 RTU/well for 2 hours at 1600xg at 16°C, leaving the vectors incubating with the cultures overnight. Starting from 24 hours post transduction onwards the cultures were selected with 5 µg/ml of blasticidin, where needed. For lentiCRISPR vectors, 0.5 RTU/well were used following the same transduction protocol and cells were selected with 0.5 µg/ml of puromycin.

When targeting genomic EGFP sequences, cells were collected and analysed using a FACSCanto flow cytometer (BD Biosciences) to quantify the percentage of EGFP loss or induction (gene substitution experiments).

Western blots. The protocol for the preparation of cell lysates and western blots has been reported in the Methods section of the first part of the results. The membranes were incubated with mouse anti-FLAG (Sigma) for detecting SpCas9 and St1Cas9, mouse anti-α-tubulin (Sigma), rabbit anti-GFP (Santa Cruz Biotechnology), mouse anti-roTag mAb³⁵⁸ and with the appropriate HRP conjugated goat anti-mouse (KPL) or goat anti-rabbit (Santa Cruz Biotechnology) secondary antibodies for ECL detection.

Detection of Cas9-induced genomic mutations. Genomic DNA was isolated at 72h post-transfection or as indicated for transduction experiments, using the DNeasy Blood & Tissue kit (Qiagen). PCR reactions to amplify genomic loci were performed using the Phusion High-Fidelity DNA polymerase (Thermo Scientific). Samples were amplified using the oligos listed in the Appendix Table 10. Purified PCR products were analysed either by sequencing and applying the TIDE tool³¹⁰ or by T7 Endonuclease 1 (T7E1) assay (New England BioLabs). In the latter case PCR amplicons were denatured and re-hybridized before digestion with T7E1 for 30 min at 37°C. The digested material was then separated using standard agarose gel electrophoresis and quantified using the ImageJ software. Indel formation was calculated according to the following equation: % gene modification = $100 \times (1 - (1 - \text{fraction cleaved})^{1/2})$.

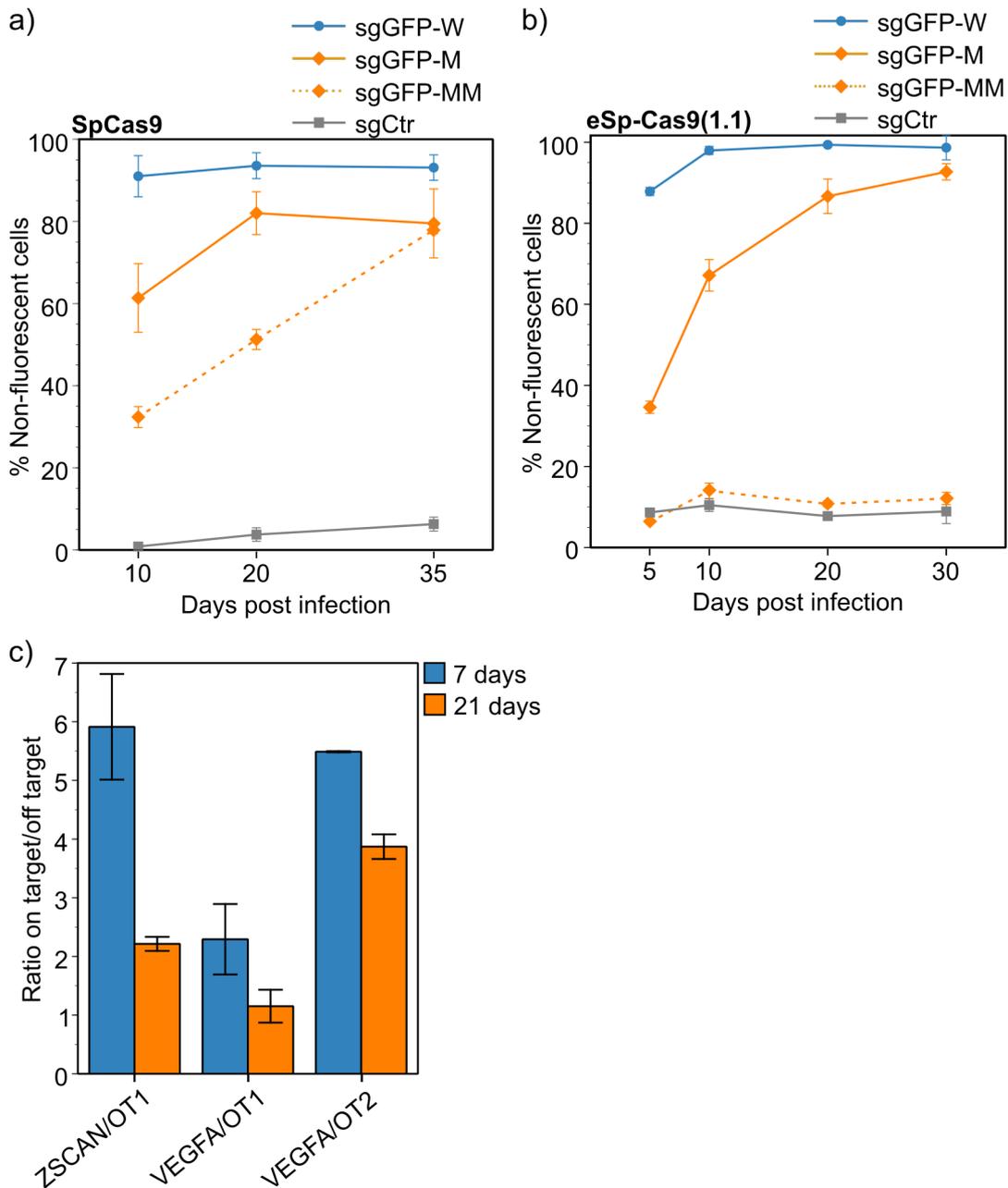
GUIDE-seq experiments. GUIDE-seq was performed as previously described with few modifications³⁰⁵. Briefly, 293T cells stably expressing SpCas9 (293T-SpCas9) were transfected with 250ng of sgRNA-encoding plasmid and 10pmol of annealed GUIDE-seq oligonucleotides (dsODNs) using Lipofectamine 3000 transfection reagent (Invitrogen). Four days post transfection genomic DNA was extracted using the DNeasy Blood and Tissue kit (Qiagen). Library preparation and data analysis have been described in detail in the Methods section of the first part of the Results.

Results

Effects of long-term Cas9 expression on editing activity

We first wanted to evaluate the impact on off-target activity produced by long term expression of SpCas9 into cells. To obtain stable Cas9 expression, we transduced the 293-iEGFP cell line, carrying a single chromosomal copy of EGFP, with a lentiviral vector expressing SpCas9 together with sgRNAs that can fully (sgGFP-W, for its ability to target the wild-type sequence) or partially (sgGFP-M or sgGFP-MM, for the presence of a single or a double mismatch in the spacer sequence, respectively; see **Appendix Table 8**) anneal to the same EGFP target. The tolerance of SpCas9 for single (sgGFP-M) or double (sgGFP-MM) mismatches in cleaving EGFP allows for the quantification of the nuclease specificity. While the percentage of EGFP negative cells obtained with the on-target sgRNA quickly reached a plateau at 10 days post-infection, the two mismatched sgRNAs generated unspecific EGFP knockout which accumulated over time (**Fig. 29a**). The delivery of the recently developed more specific eSpCas9(1.1) variant³⁴⁸ guided by the same sgRNAs only partially reverses the time dependent accumulation of off-target cleavages (**Fig. 29b**). The slight increase in EGFP-negative cells over time in the control samples is most likely caused by the progressive loss of functionality of the Tet-mediated inducible expression circuit, which is probably due to cellular stress. Consistently, the analysis of the on/off ratios calculated from the editing of two genomic loci (ZSCAN and VEGFA) and their related off-target sites after long-term Cas9 expression indicated a progressive decrease over time, thus confirming increased off-target cleavages (**Fig. 29c**). These results clearly show that the delivery of SpCas9 through a conventional lentiviral system correlates with increased off-target activity and this is particularly evident over time due to prolonged SpCas9 expression.

Figure 29 (next page). Long term expression of Cas9 correlates with the accumulation of off-target cleavages. (a) Time course curves of the percentages of 293-iEGFP non-fluorescent cells obtained after transduction with a lentiviral vector (lentiCRISPR) expressing SpCas9 together with either a perfectly matching sgRNA (sgGFP-W) or two different sgRNAs containing one or two mismatches with the target sequence (sgGFP-M and -MM, respectively). A vector expressing an irrelevant sgRNA was used as control (sgCtr). (b) As in (a) using a lentiviral vector expressing a SpCas9 variant with increased fidelity (eSpCas9(1.1)). (c) DNA modification specificity, defined as on-target/off-target indels frequency ratio, after long term SpCas9 expression with sgRNAs targeting the VEGFA and ZSCAN endogenous loci. Percent modification of previously validated off-target sites was quantified by TIDE analysis after one week and 21 days post-transduction. For all the experiments, cells were selected with puromycin in order to eliminate non-transduced cells. In panels (a-c) data presented as mean \pm s.e.m. for n=2 independent experiments.



Design of a Self-Limiting Cas9 circuitry for Enhanced Safety (SLiCES)

To generate a transient SpCas9 activity peak in target cells in order to allow the modification of the intended target with the parallel switch-off of nuclease expression, we developed a Self-Limiting Cas9 circuitry for Enhanced Safety and specificity (SLiCES) (schematized in **Fig. 30a**). Cas9 was expressed in target cells together with two different sgRNAs: one guide molecule targets Cas9 activity towards a genomic locus of interest; in parallel, the second sgRNA allows the recognition of the Cas9 coding sequence itself, introducing indel mutations and thereby shutting down nuclease expression. The presence of a second sgRNA could in principle increase the likelihood of introducing a

higher number of unwanted cleavages into the cellular genome. However, given the artificial nature of the nucleotidic sequence encoding Cas9, derived from the human codon optimization of a prokaryotic gene, no highly homologous sequences were expected to occur in the host genome. This allowed to easily select guide RNAs with minimal off-target sites into the human genome and particularly into genes, according to available online prediction tools. The self-limiting SpCas9 circuitry was set up by selecting three different sgRNAs targeting different regions of the SpCas9 coding sequence (sgCas-a, -b and -c). The potential off-target sites generated by sgCas-a, -b and -c were evaluated in 293T cells stably expressing SpCas9 through GUIDE-seq analysis, a genome-wide unbiased approach. Both sgCas-a and -c did not generate detectable off-target while retaining the ability to efficiently cleave the on-target site. On the other hand, sgCas-b produced only 6 off-target sites (**Fig. 30b**).

When co-expressed with SpCas9, all three guide molecules were shown to efficiently downregulate SpCas9 levels (**Fig. 30c** and **30d**, upper panels). Co-transfection in EGFP expressing cells of any of the three self-targeting sgRNAs (sgCas-a, -b or -c) together with SpCas9 and a sgRNA that fully base pairs with the EGFP target sequence (sgGFP-W) reduced intracellular EGFP to levels (4-10% of residual protein) similar to those detected in cells co-transfected with the same sgGFP-W and a control sgRNA (sgCtr), where no downregulation of SpCas9 was induced (**Fig. 30c**). These results demonstrated that the DNA editing activity is not impaired when SpCas9 is inactivated through the SLICES circuit. A similar experiment performed using a sgRNA targeting EGFP with a single mismatch within the seed region at the last nucleotide before the PAM sequence (sgGFP-M) showed non-specific EGFP downregulation, with almost 60% decrease of EGFP intracellular levels, in accordance with the long-term experiments presented in the previous section. This effect was less pronounced (~25-40% reduction) in cells where SpCas9 expression was downregulated through the self-limiting Cas9 circuit (sgCas-a, -b). Of note, the sgCas-c guide RNA was not able to reduce unspecific EGFP knockout, even though intracellular Cas9 levels were indeed diminished (**Fig. 30c**). The different levels of non-specific EGFP downregulation closely reflected the ability of individual sgRNAs to decrease SpCas9 intracellular levels: sgCas-a, which generated the lowest non-specific EGFP downregulation (73% residual EGFP, **Fig. 30c**), showed the highest SpCas9 disruption activity (**Fig. 30c**, upper panel). Similar results

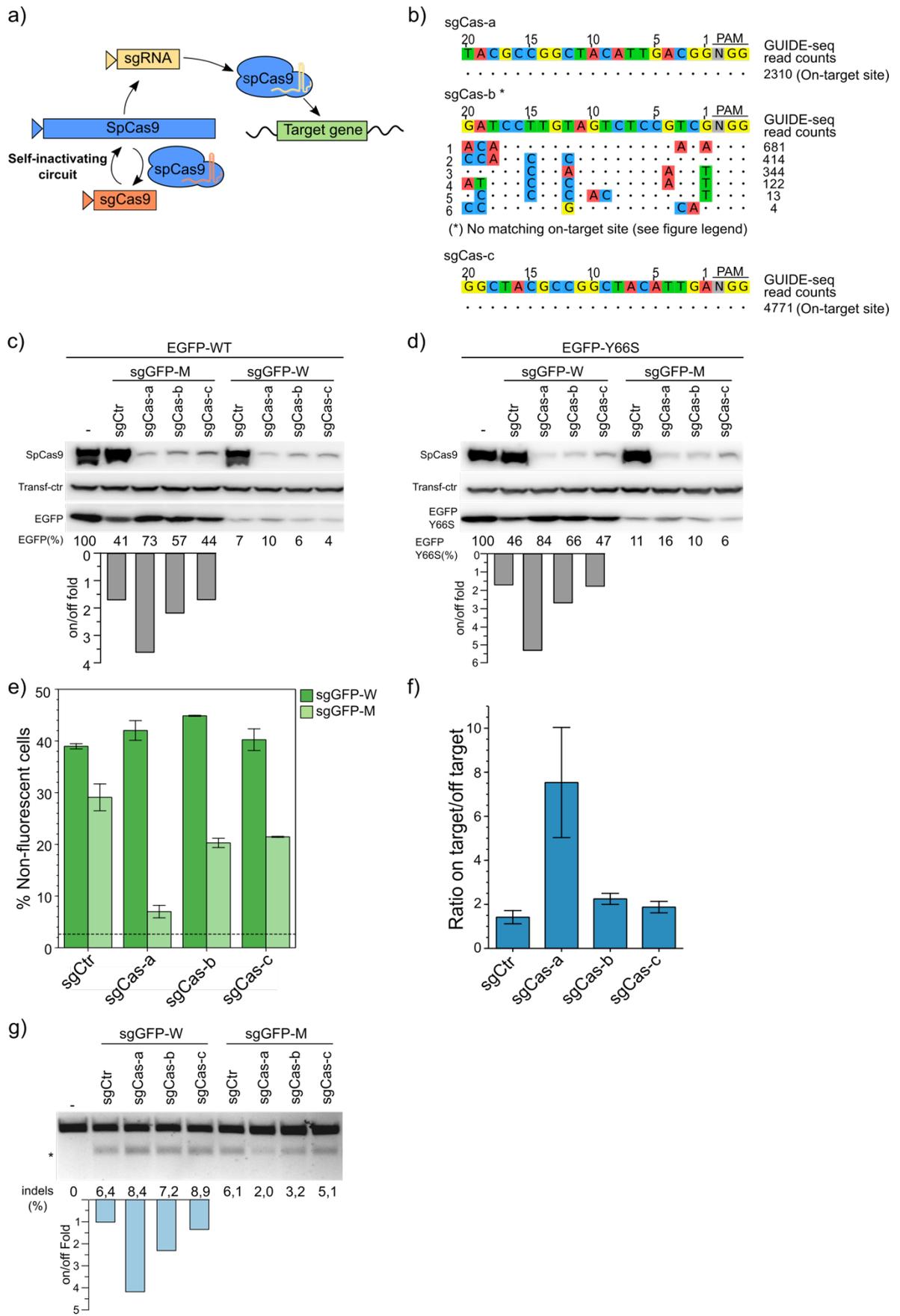


Figure 30 (on previous page). The SLiCES circuit. (a) Scheme of the SLiCES circuit. SpCas9 is expressed together with sgRNAs directed to its own open reading frame (ORF) for self-limiting activity and to a selected target sequence. **(b)** GUIDE-seq analysis performed in 293T cells stably expressing SpCas9 and transfected with sgCas-a, -b and -c individually. Genomic DNAs obtained from three independent experiments were pooled prior to library preparation. (*) The on-target site was not detected in sgCas-b samples since this guide targets the 3xFLAG tag fused to SpCas9 which was absent from SpCas9 expressed in 293T-Cas9 cells. **(c)** Regulation of SpCas9 and EGFP target gene expression by the SLiCES circuit. Western blot analysis of 293T cells co-transfected with plasmids expressing EGFP, SpCas9 and sgRNAs fully (sgGFP-W) or partially matching (sgGFP-M) the EGFP coding sequence in combination with three sgRNAs targeting the SpCas9 ORF (sgCas-a, -b, -c) or a control sgRNA (sgCtr), as indicated. **(d)** Western blot of cells co-transfected with plasmids expressing EGFP-Y66S, SpCas9, sgRNAs perfectly matching (sgGFP-M) or containing one mismatch (sgGFP-W) with the EGFP-Y66S target sequence together with sgRNAs specific for the SpCas9 ORF (sgCas-a, -b, -c) or a control sgRNA (sgCtr), as indicated. For (c) and (d): Lower graph reports the ratio of the percentages of decreased EGFP or EGFP-Y66S levels obtained using on-target sgRNAs over the percentages obtained with the off-target sgRNAs in the presence of sgCas-a, -b, -c, as indicated. Lane (-) corresponds to a reference sample containing the non-targeting sgCtr only; transfection efficiency was normalized using roTag tagged MHC-1 α expression plasmid (Transf-ctr); SpCas9 was detected using an anti-FLAG antibody. **(e)** Percentage of non-fluorescent 293-iEGFP cells obtained after expression of different self-limiting SpCas9 circuits. Cells were transfected with sgRNAs perfectly matching (sgGFP-W) or containing one mismatch (sgGFP-M) with the EGFP ORF together with three sgRNAs targeting the SpCas9 ORF (sgCas-a, -b, -c) or a control sgRNA (sgCtr), as indicated. The dashed line represents the average background of EGFP negative cells. **(f)** On/off ratios were calculated from the percentage of EGFP negative cells obtained in (b) with sgGFP-W (on-target) relative to sgGFP-M (off-target) in combination with different SLiCES circuits (sgCas-a, -b or -c) or a non-targeting (sgCtr) sgRNA, as indicated in the graph. **(g)** Representative T7 Endonuclease assay from cells expressing different SLiCES circuits. The on/off specificity ratio was calculated by measuring indels formation in the EGFP gene in the presence of sgGFP-W or sgGFP-M together with a control sgRNA or the three sgRNAs targeting the SpCas9 ORF (sgCas-a, -b, -c). Lane (-) corresponds to a reference sample containing the non-targeting sgCtr only. (*) Indicates the expected band obtained by T7 endonuclease activity. Error bars in (e) and (f) represent s.e.m. for n=2 independent experiments.

were obtained with a reciprocal experiment where cells were transiently transfected with a mutated EGFP target characterized by a single nucleotide substitution (EGFP-Y66S) that fully matched the sgGFP-M sequence and conversely differed from sgGFP-W by a single nucleotide immediately adjacent to the PAM sequence (**Fig. 30d**). In both experimental setups the introduction of SLiCES induced a 2 to 4-fold improvement in target specificity, as defined by the ratio between SpCas9 editing in cells targeted by the perfectly matched sgRNA over the mismatched sgRNA (**Figs. 30c-d**, lower panel). This result was also confirmed in 293-iEGFP cells stably expressing a single copy of the EGFP gene, where we obtained a 4 to 5-fold improvement of the on-/off-ratio as measured both by flow cytometry to quantify the number of EGFP-negative cells and by T7 endonuclease I assay to assess indel formation at the genomic EGFP target locus (**Figs. 30e-g**).

Previous reports^{268,361} have demonstrated that it is possible to modulate SpCas9 cleavage activity against its targets by altering the structure of the constant portion of the guide RNA molecule. To test whether the optimization of the guide RNAs may

further improve the on-target specificity, we generated alternative SLICES circuits where the sgRNAs were structurally modified to increase their transcription and their interaction with SpCas9, according to previous studies²⁶⁸. The optimization of the SpCas9-targeting sgCas-a guide RNA (sgCas-a-opt), which enhances the efficiency of nuclease removal as demonstrated by western blot analysis (**Fig. 31d**), produced a 9-fold improvement in cleavage specificity when tested in a stable EGFP-expressing cell

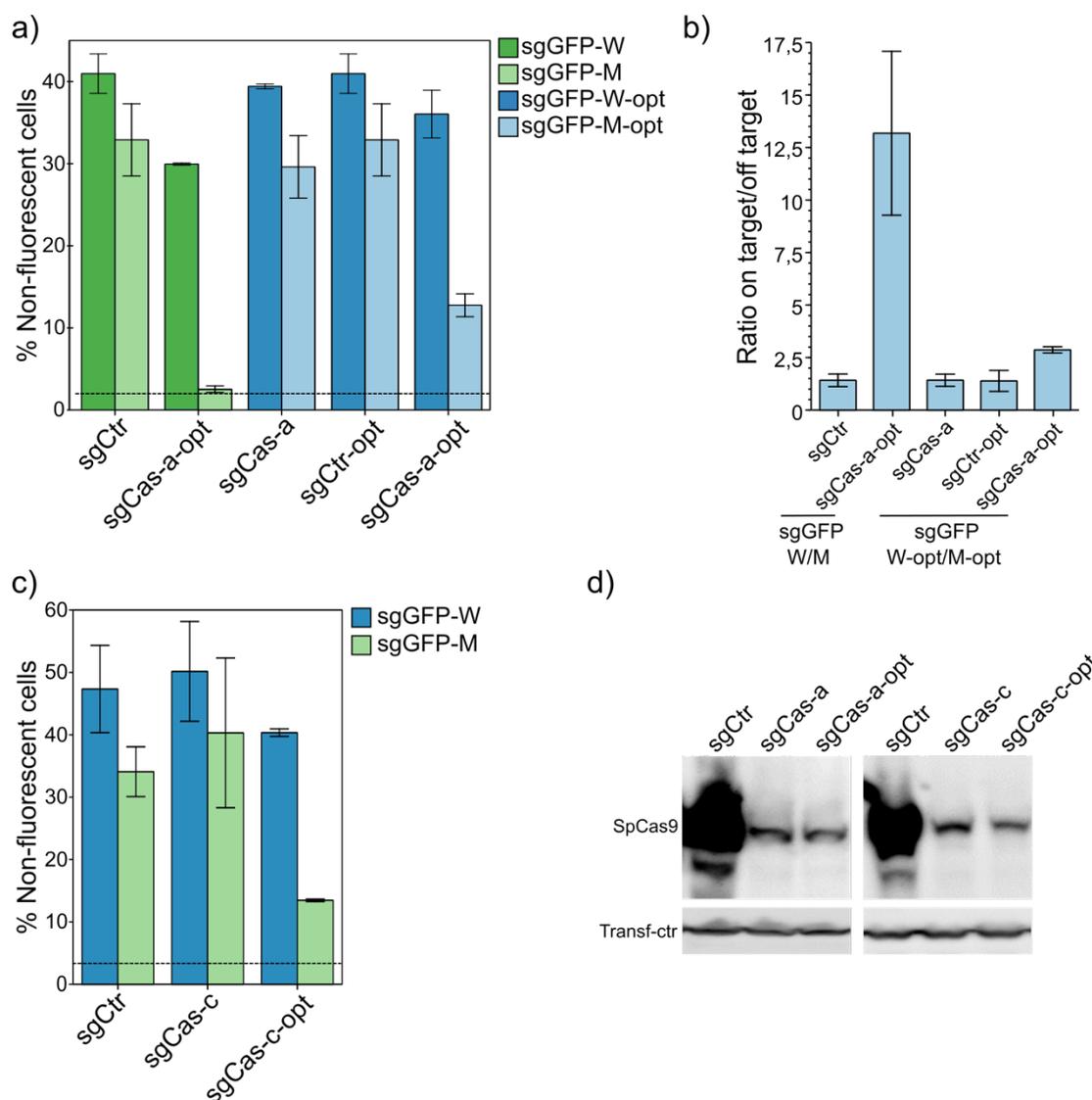


Figure 31 (on previous page). Effect of sgRNAs optimization on SLICES circuit. (a) Percentage of non-fluorescent 293-iEGFP cells obtained after transfection of SpCas9 with sgRNAs targeting EGFP (sgGFP-W or sgGFP-W-opt, if optimized) or containing a single mismatch (sgGFP-M or sgGFP-M-opt, if optimized) together with the sgCas-a. The optimized version of the SLICES sgRNA (sgCas-a-opt) was tested with both standard and optimized sgRNAs targeting EGFP, as indicated. **(b)** On/off ratios were obtained from the percentage of EGFP negative cells after targeting a single chromosomal EGFP gene copy (293-iEGFP cells) with sgGFP-W or sgGFP-W-opt (on-target) relative to sgGFP-M or sgGFP-M-opt (off-target) in combination with the self-targeting sgCas-a, its optimized variant, or a non-targeting (sgCtrl) sgRNA, as indicated in the graph. **(c)** Percentage of non-fluorescent 293-iEGFP cells obtained after transfection of SpCas9 with sgRNAs targeting EGFP (sgGFP-W) or containing a single mismatch (sgGFP-M) together with the sgCas-c or sgCas-c-opt, if optimized. **(d)** Western blot analysis of 293T cells co-transfected with SpCas9 and sgCas-a or sgCas-a-opt and sgCas-c or sgCas-c-opt. SpCas9 was detected using an anti-FLAG antibody. Transfection efficiency was normalized using roTag tagged MHC-Ia expression plasmid (Transf-ctr). Error bars represent s.e.m. for n=2 independent experiments.

line, further ameliorating the effect observed with conventional guides (**Figs. 31a-b**). Consistently, the optimization of the least active self-inactivating sgRNA (sgCas-c-opt) resulted in reduced off-target activity paralleled by a further decrease in SpCas9 intracellular levels (**Figs. 31c-d**). Conversely, the optimization of the sgRNA towards the intended target site (sgGFP-W-opt and sgGFP-M-opt) did not increase the specificity in combination either with sgCas9-a or sgCas9-a-opt and, on the contrary, produced a reduction in the on-/off-target ratio when compared to the standard SLiCES circuit (**Figs. 31a-b**). Presumably, the enhanced downregulation of EGFP driven by the sgGFP-W-opt correlated also with increased off-target cleavages induced by the sgGFP-M-opt sgRNA that could not be counteracted by a sufficiently rapid SpCas9 downregulation mediated by both the standard and the optimized versions of the self-limiting sgCas-a sgRNA (**Figs. 31a-b**). In conclusion, the SLiCES circuitry produced the highest on-target specificity when composed of an optimized self-limiting sgRNA (sgCas-a-opt), efficiently downregulating SpCas9, in combination with a non-optimized sgRNA targeting the site of interest (sgGFP-W/M).

Effect of SLiCES on gene editing by homologous recombination

We next evaluated the effect of the application of SLiCES on the specificity of gene targeting obtained by homologous recombination. This is particularly relevant in situations where the sequence that has to be edited shares homologies with other regions of the genome, for example when only one of the two heterozygous alleles of a gene needs to be modified. In addition, the removal of Cas9 soon after the cleavage of the target site lowers the possibility of the introduction of further indel mutations after homologous recombination with the donor template has occurred, in all the instances where the donor sequence cannot be modified to avoid re-editing by Cas9. To this aim, we used as an experimental model cells carrying a single chromosomal copy of a non-fluorescent EGFP, containing a single nucleotide substitution that impairs fluorophore maturation (Y66S). In these cells, 293-iY66S, the recovery of EGFP fluorescence following the substitution of the mutated gene with a wild-type allele was used to measure SpCas9-mediated homology-directed repair in the presence of a co-transfected donor plasmid carrying a non-fluorescent fragment of wild-type EGFP. We used the sgGFP-M (fully matching to the EGFP Y66S mutant sequence) or the sgGFP-W

(containing a single mismatch with the target) sgRNAs to introduce DSBs at the target locus. Compared to the conventional SpCas9 approach (sgCtr), the target specificity for EGFP homology-directed repair was improved by 4-fold when using the SLICES circuitry driven by the sgCas-a self-targeting guide, as measured by the on/off ratio calculated on the percentage of EGFP positive cells obtained after successful gene editing (**Figs. 32a-b**). As previously observed, the sgCas-b and sgCas-c sgRNAs were unable to increase significantly the discrimination between the matching and the mismatched EGFP-targeting guide. Additionally, further improvement (7,5-fold) was obtained with the optimized version of sgCas-a (sgCas-a-opt) (**Figs. 32a-b**), in accordance with previous EGFP knockout experiments.

Altogether, given its superior ability to modulate SpCas9 off-target activity, we decided to exploit the sgCas-a-opt self-targeting sgRNAs for further implementations of the SLICES circuit. Nevertheless, the other self-limiting guides, as well as other specifically tailored sgRNAs, could be still employed to obtain different circuit behaviours by changing the kinetics of SpCas9 downregulation into cells.

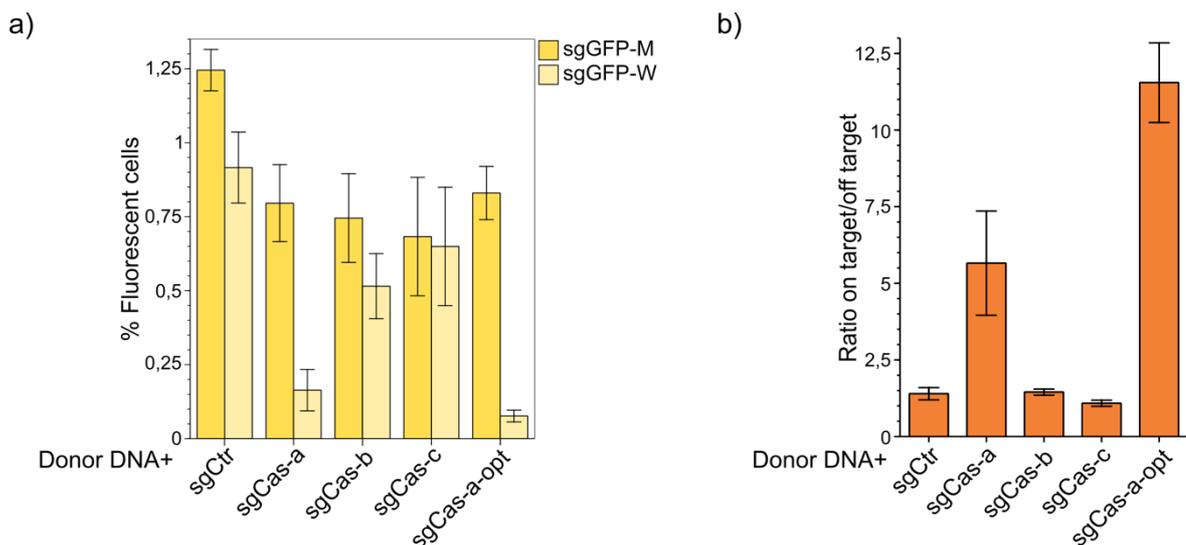


Figure 32. Specificity of homology-directed repair mediated by SLICES. (a) Percentage of fluorescent 293-iY66S cells obtained after transfection with a donor DNA plasmid (carrying a non-fluorescent fragment of wt-EGFP), SpCas9 together with sgRNAs matching (sgGFP-M) or containing one mismatch with the EGFP-Y66S target sequence (sgGFP-W) and the three sgRNAs targeting the SpCas9 ORF (sgCas-a, -b, -c or sgCas-a-opt) or a control sgRNA (sgCtr), as indicated. Homology-directed repair in the absence of sgGFP-M or sgGFP-W was about 0.01%. **(b)** Target specificity of SpCas9 activity expressed as on/off ratios using different self-limiting circuits applied to a gene substitution model. On/off ratios were obtained from the percentage of EGFP positive cells generated by SpCas9 homology-directed repair of the EGFP-Y66S mutation with the sgGFP-M (on-target) relative to the sgGFP-W (off-target) sgRNAs in combination with a DNA donor plasmid (carrying wild-type EGFP sequence) in 293-iY66S cells containing a single mutated EGFP gene copy. Error bars represent s.e.m. for n=2 independent experiments.

Adaptation of SLiCES to other RNA-guided nucleases

To demonstrate that the SLiCES methodology is readily transferrable to other RNA-guided nucleases, this self-limiting approach was adapted to a Cas9 orthologue deriving from *Streptococcus thermophilus* CRISPR1 locus (St1Cas9). We selected three self-targeting sgRNAs (sgCas-St1-1, -2 and -3) and verified that they were able to induce St1Cas9 downregulation (**Fig. 33b**) when co-expressed with St1Cas9. To test their effect on St1Cas9 specificity we developed a reporter plasmid that allows to easily detect cleavage events by measuring NHEJ-dependent expression of EGFP. This NHEJ-reporter contains a target sequence recognized by the sgRNA of interest inserted between an SV5 tag and the EGFP coding sequences, with the EGFP ORF positioned out of frame with respect to the starting ATG codon of the SV5 tag ORF. A stop codon has been added to the SV5 frame, immediately after the target sequence, to stop its translation. After Cas9-mediated cleavage of the target sequence and repair by NHEJ, indel mutations are inserted randomly at the breakpoint, inducing the shift of the EGFP ORF in the same frame of the SV5 tag ORF, allowing its detection either directly by fluorescence measurements or by western blot through the SV5-tag. If no editing occurs, the reporter plasmid expresses only an SV5 peptide of few kDa. A schematic representation of the NHEJ reporter is presented in **Fig. 33a**. We generated alternative versions of the NHEJ-reporter containing two target sequences differing for a single nucleotide (NHEJ-Rep.W and NHEJ-Rep.M). We next co-transfected 293T cells with St1Cas9, an sgRNA perfectly matching the NHEJ-Rep.W together with each self-targeting sgRNA or a control guide and the NHEJ-Rep.W or the NHEJ-Rep.M reporters. Reporter activation was then measured by western blot revealing efficient on-target cleavage (activation of the NHEJ-Rep.W) when St1Cas9 was expressed with both sgCas-St1-1 and sgCas-St1-2 sgRNAs, while less activity was detected in the presence of the sgCas-St1-3 self-targeting guide RNA (**Fig. 33b**). When off-target activity was assessed by quantifying NHEJ-Rep.M activation, we observed increased discrimination of the off-target site in the presence of the SLiCES circuit driven by sgCas-St1-1, while no relevant effect was obtained with the other two self-targeting guide RNAs, as demonstrated by the on/off ratios calculated from the densitometric analysis of western blot images (**Figs. 33b-c**). These data indicate that the SLiCES technology can be easily adapted to other RNA-guided nucleases without major modification to circuit design, provided

that a self-targeting guide RNA with the desired characteristics can be identified.

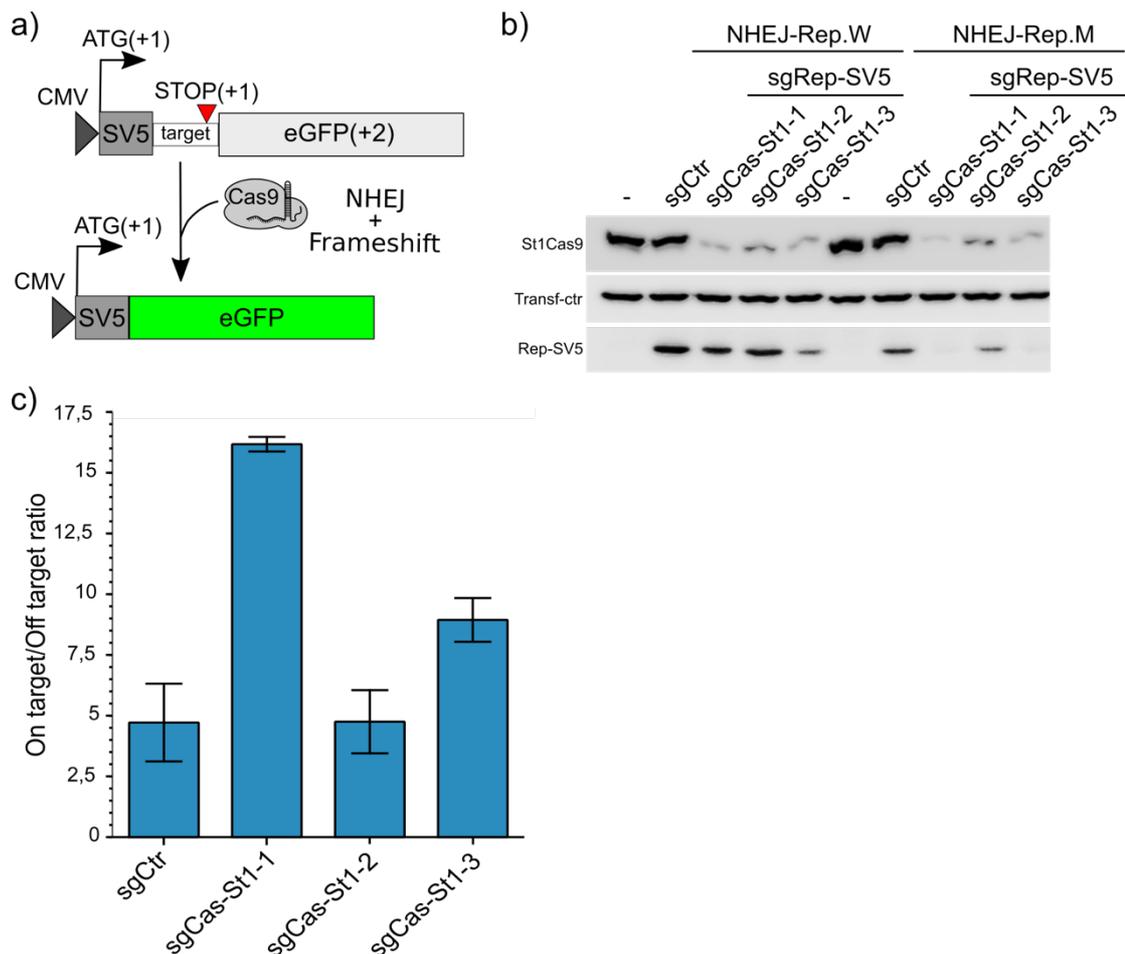


Figure 33. Activity of SLiCES with *Streptococcus thermophilus* CRISPR1/Cas9. (a) Schematic representation of the SV5-GFP-based NHEJ reporter. (b) Evaluation of St1Cas9 activity expressed through the SLiCES system. Western blot of 293T cells transfected with St1Cas9, the NHEJ reporter carrying either a target sequence that fully base pairs with the sgRep-SV5 (NHEJ-Rep.W) or including one mismatch (NHEJ-Rep.M), the sgRNA sgRep-SV5 and three different St1Cas9 targeting sgRNAs (sgCas- St1, -2, -3). St1Cas9 mediated cleavages are detected by frameshift of the EGFP ORF and SV5-EGFP expression by the NHEJ reporter. Lane (sgCtr) corresponds to a sample transfected with a non-self-targeting sgRNAs; lane (-) corresponds to a sample transfected with a non-targeting sgRNA. St1Cas9 was detected using an anti-FLAG antibody. Western blot is representative of n=2 independent experiments. (c) Modulation of St1Cas9 expression by self-limiting circuits increases on target specificity. On/off target ratios calculated from levels of SV5-EGFP expression obtained from cells transfected with NHEJ- Rep.W or NHEJ-Rep.M together with sgRep-SV5 in combination with St1Cas9 targeting sgRNAs (sgCas-St1, -2, -3) or a non-self-targeting sgRNAs sgCtr, as in (b). Error bars represent s.e.m. for n=2 independent experiments.

SLiCES specificity towards endogenous genomic loci

Next, the target specificity of SpCas9 and of the most effective version of the SLiCES circuit (driven by sgCas-a-opt) were comparatively analysed. After transfection of the SLiCES components into 293T cells, four genomic sites (VEGFA site 3, ZSCAN2 and two targets in the EMX1 locus, EMX1-K and EMX1-R) and two previously validated off target sites³⁴⁷ for each sgRNA were analysed by Tracking of Indels by DEcomposition (TIDE)³¹⁰

revealing that the SLiCES approach increased cleavage specificity by approximately 1.5-2.5 fold, as measured by the fold improvement in the on/off ratio, without significantly diminishing on-target cleavage efficiency (**Figs. 34a-d**). These results, in accordance with previous data obtained on both transient and stable EGFP experimental models, demonstrated the efficiency of SLiCES in reducing SpCas9 off-target activity.

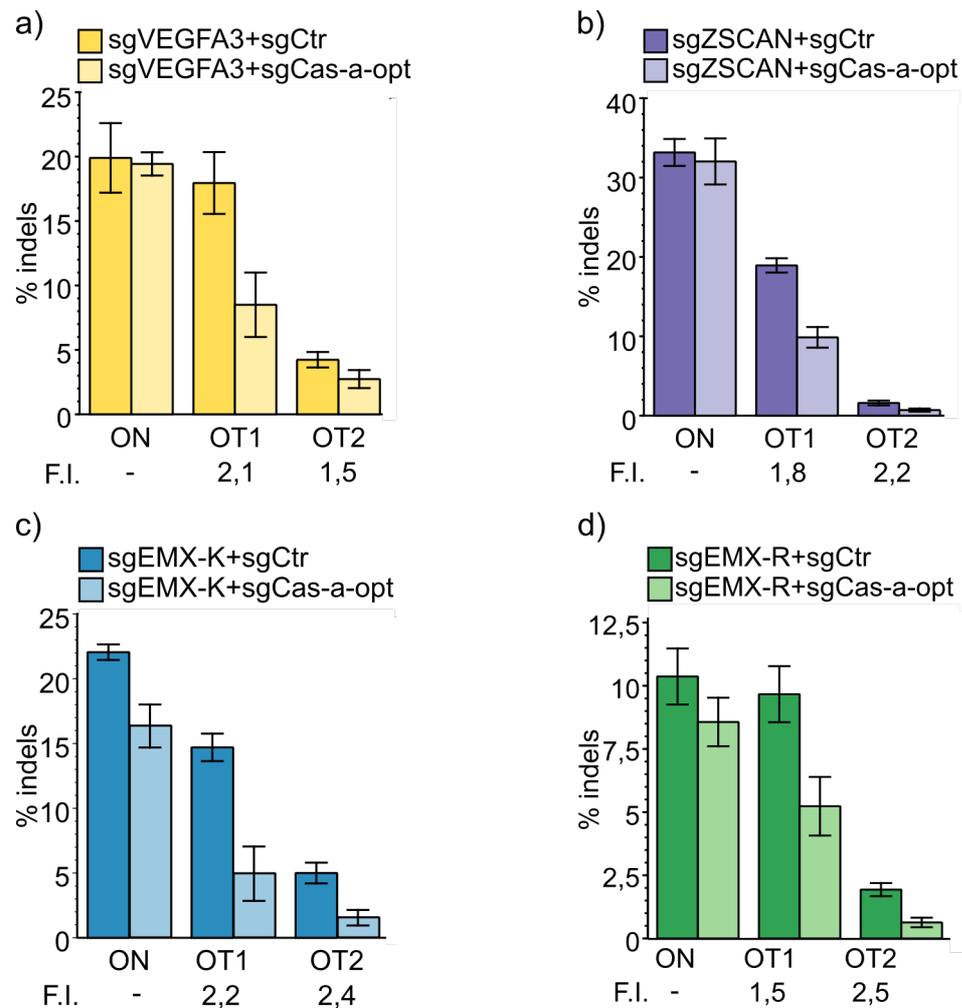


Figure 34. Activity of SLiCES on endogenous loci. Indels formation induced by the SLiCES circuit (sgCas-a-opt) targeting the VEGFA site 3, ZSCAN2, EMX1 loci and their respective validated off-target sites. Fold increase (F.I.) of the on/off ratio with the sgCas-a-opt relative to the sgCtr is reported below the graphs for each off-target. Percent modification was quantified by TIDE analysis. Error bars represent s.e.m. for n=2 independent experiments.

Adaptation of SLiCES to lentiviral delivery

Lentiviral vectors are optimal tools for the delivery of transgenes by enabling the transduction of difficult-to-transfect cell lines and primary cells as well as for their *in vivo* applicability. The self-limiting SpCas9/sgRNA circuitry regulated by the best selected self-limiting sgRNA (sgCas-a-opt) was thus transferred to a lentiviral system to

generate lentiSLiCES (**Figs. 35a-b**). Two technical expedients had to be adopted in order to implement an all-in-one lentiviral system to deliver SLiCES. First of all, to avoid the leaky expression of SpCas9, and the consequent degradation of DNA during plasmid preparation in bacteria, an intron was introduced into the SpCas9 open reading frame to form an expression cassette divided in two exons (exon 1 and 2, schematized in **Fig. 35a**). As splicing does not occur in bacteria, the transcripts produced are translated as a catalytically inactive SpCas9 fragment. Second, to circumvent the self-cleavage activity during lentiviral vector production, Tetracycline inducible (TetO) promoters were introduced to regulate both SpCas9 and the self-targeting sgRNA expression. The TetO promoter is negatively regulated by a specific repressor, TetR, which is expressed in producing cells and, in the absence of doxycycline, inhibits transcription through its binding to tetracycline operator sequences located within the SpCas9 and sgRNA promoter regions (schematized in **Fig. 35b**).

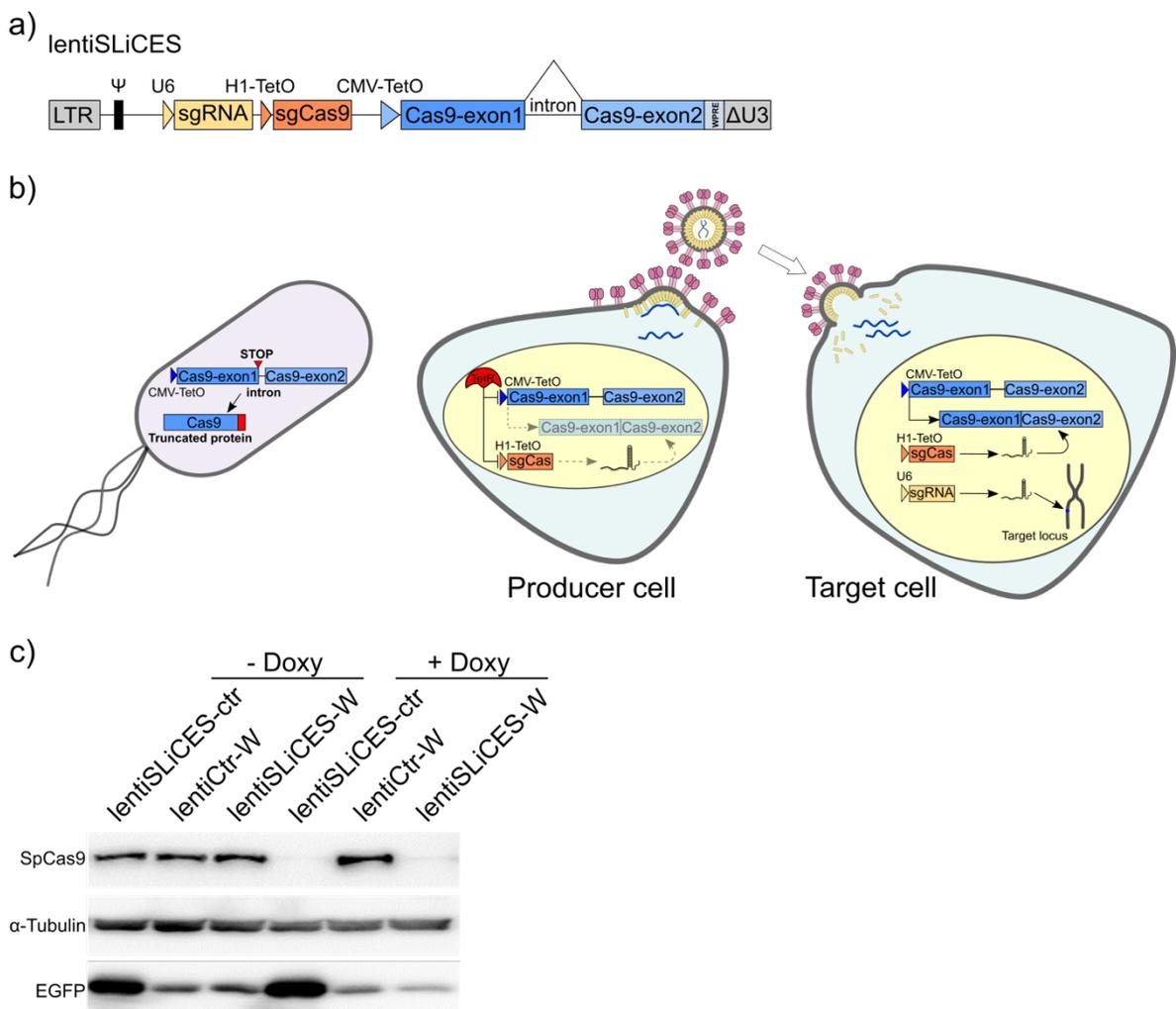


Figure 35 (on previous page). lentiSLiCES. (a) Graphical representation of lentiSLiCES viral vector. **(b)** Steps required for the production of the lentiSLiCES viral vectors. SpCas9 expression is prevented in bacterial cells to allow plasmid amplification through the introduction of a mammalian intron within the SpCas9 open reading frame. Production of lentiSLiCES viral particles is obtained in cells stably expressing the Tetracycline Repressor (TetR) to prevent SpCas9 and sgCas self-limiting sgRNA expression driven by Tet repressible promoters. In target cells the absence of the TetR allows the expression of the lentiSLiCES circuit leading to target genome editing and simultaneous SpCas9 downregulation. **(c)** lentiSLiCES circuit behaviour in viral vector packaging cells. Western blot analysis of 293TR cells transfected with EGFP and self-limiting or non- self-limiting transfer vectors carrying sgGFP-W (lentiSLiCES-W or lentiCtr-W, respectively), or with lentiSLiCES carrying a non-targeting sgRNA (lentiSLiCES-Ctr). Cultures were treated as indicated with doxycycline to upregulate expression of SpCas9 and of the self-targeting sgCas-a-opt. SpCas9 was detected using an anti-FLAG antibody. Western blot is representative of n=2 independent experiments.

To verify the efficacy of this failsafe mechanism, we analysed producer cell lysates during vector production. Even in the absence of doxycycline, we detected low levels of Cas9 expression likely connected to the translation of spliced viral genomes, generated by LTR-driven unrepressed transcription (**Fig. 35c**). Notably, no reduction in SpCas9 intracellular levels was observed in the presence of the sgCas-a-opt self-targeting sgRNA, indicating that the additional repression of the self-targeting guide promoter was effectively blocking SpCas9 self-editing (**Fig. 35c**). On the other hand, the drop of SpCas9 intracellular levels in producing cells observed with the activation of the self-limiting circuitry by the addition of doxycycline to the culture medium demonstrates the strict requirement of the repressible promoters at viral production steps in order to obtain un-altered lentiSLiCES particles (**Fig. 35c**).

Characterization of lentiSLiCES specificity

To evaluate the on/off target activity of lentiSLiCES, we followed the percentage of EGFP negative 293-multiEGFP cells at different time points after transduction with self-limiting lentiviral vectors either carrying the specific sgRNA sgGFP-W (lentiSLiCES-W) or the mismatched sgGFP-M (lentiSLiCES-M). We then compared these measurements with the effect obtained with corresponding non-self-limiting lentiviral vectors expressing the same sgRNAs directed towards EGFP (lentiCtr-W or -M). Both lentiCtr-W and lentiSLiCES-W showed similarly stable on-target activity at all the time points within a 3 weeks period (**Fig. 36a**). Conversely, the percentage of EGFP cells unspecifically targeted by the sgGFP-M increased in time with the lentiCtr delivery system; this effect was not observed when the same was sgRNA delivered through lentiSLiCES throughout the 3 weeks period (**Fig. 36a**). Therefore, lentiSLiCES generated no off-target accumulation in time (compare day 7 and day 21, **Fig. 36a**). Consistently,

at the end-point we observed the largest difference between the ratios of the EGFP negative cells obtained with the sgGFP-W over the sgGFP-M delivered either through the lentiSLiCES (on/off ratio ~5) or the lentiCtr systems (on/off ratio ~2) (**Fig. 36b**).

We then evaluated the effect of lentiSLiCES on the editing of the VEGFA site 3 and ZSCAN2 genomic loci and related off-target sites. In agreement with previous results on the EGFP locus, the target specificity of lentiSLiCES towards these endogenous sequences showed significant improvement as compared to the non-self-limiting lentiCtr (approximately 2-4 fold increase when considering the on/off-ratios) (**Fig. 36c-d**). Of note, we also observed a reduction in the cleavage efficiency of the on-target sites that was more pronounced for the ZSCAN2 locus and was most likely connected with the transient nature of SpCas9 expression induced by the self-limiting circuit (**Fig. 36c**). These data suggest that the decreased expression of SpCas9 obtained through the SLiCES circuit improves editing specificity. Indeed, at early time points (2 days post-transduction) the SpCas9 protein was already much less present in cells treated with the lentiSLiCES than in cells treated with the non-self-limiting lentiviral control (lentiCtr) (**Fig. 36e**). Notably, in lentiCtr treated cells the levels of SpCas9 remained stable and higher than in lentiSLiCES transduced cells, where no nuclease could be detected at any later time point. To functionally assess the level of SpCas9 activity delivered through lentiSLiCES, the NHEJ reporter plasmid (NHEJ-Rep.W) expressing an SV5 tag fused with EGFP (SV5-EGFP) upon targeted nuclease activity (schematized in **Fig. 33a**) was transfected in cells transduced with lentiSLiCES or lentiCtr non self-limiting vectors both expressing a guide RNA that was able to recognize the target sequence inserted in the reporter. Activation of the reporter thus correlates with the residual SpCas9 activity. The NHEJ-Rep.W revealed that SpCas9 delivered through the lentiCtr was active at all time points following transduction, as expected, while the activity of SpCas9 carried by the lentiSLiCES was detected 2 days after transduction, but could not be observed at later time points (30 days), indicating correct disruption of the SpCas9 coding sequence by the self-targeting sgRNA (**Fig. 36f**).

Altogether these data demonstrate the efficient adaptation of the SLiCES circuit to an all-in-one lentiviral vehicle that after transduction efficiently edits the intended locus while reducing off-target cleavages and promoting a parallel switch-off mechanism, leaving no long-term trace of SpCas9 activity in transduced cells.

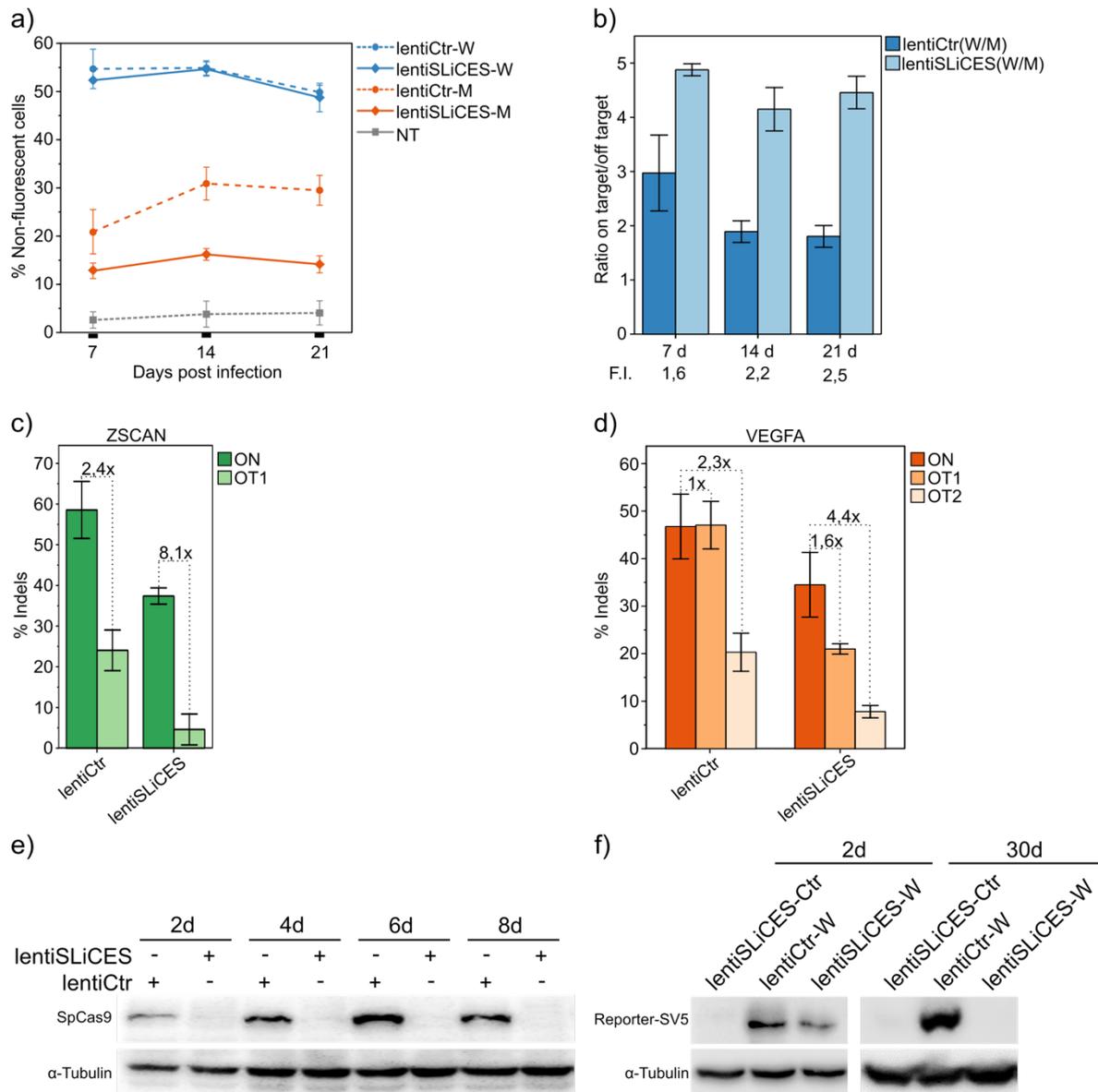


Figure 36. Genome editing with lentiSLiCES vectors. (a) EGFP knock-down by lentiSLiCES vectors. Time course curves of the percentages of EGFP negative 293-multiEGFP cells, following transduction with lentiviral vector carrying self-targeting (lentiSLiCES) or non-self-targeting (lentiCtr) sgRNAs in combination with either sgGFP-W (on-target) or sgGFP-M (off-target) sgRNAs, as indicated in the graph. Error bars represent s.e.m. for n=2 independent experiments. **(b)** Target specificity of SpCas9 delivered through the lentiSLiCES. On/off ratios were calculated from the percentages of EGFP negative cells reported in (a). Below the graphs is reported the fold increase (F.I.) of specificity calculated from the on/off ratios at each time point. Error bars represent s.e.m. for n=2 independent experiments. **(c-d)** Indel formation induced by lentiSLiCES vectors at the ZSCAN and VEGFA loci and at their validated off-target sites. Percent modification was quantified by TIDE analysis on genomic DNA collected 20 days post-transduction and selection with blasticidin. Values indicate the on/off ratios calculated from indels obtained with each off target. Error bars represent s.e.m. for n=2 independent experiments. **(e)** Expression levels of SpCas9 at the indicated time points after transduction with lentiSLiCES or with lentiCtr. SpCas9 was detected using an anti-FLAG antibody. Western blot is representative of n=2 independent experiments. **(f)** SpCas9 activity was monitored by SV5-EGFP protein levels produced by the NHEJ-reporter plasmid transfected in 293-multiEGFP cells one day before or 28 days after transduction with lentiSLiCES targeting EGFP (lentiSLiCES-W) or with a lentiSLiCES control vector (lentiSLiCES-Ctr). Cas9 activity was detected at 2 days or 30 days post-transduction, as indicated. The activity of the non-self-limiting lentiCtr-W vector targeting EGFP was monitored at the same time points for comparison. Western blot is representative of n=2 independent experiments.

Discussion

The limitations of SpCas9-based genome editing clearly emerge from our data showing that long term nuclease expression delivered through lentiviral systems as well as the short-term transient expression of Cas9 may result in the generation and the accumulation of unwanted cleavages in cellular genomes. This issue is particularly relevant to the field and has been widely reported in literature^{305,331,341,347,348,362,363}, raising concerns on the prospective therapeutic application of Cas9 into patients. Of note, in some instances, this detrimental effect could not be completely overcome even with the recently developed, more specific, SpCas9 variants eSpCas9(1.1)³⁴⁸ and SpCas9-HF1³⁴⁷. This limitation is even more obvious where the choice for the sgRNA is constrained by the target site. A practical real-world example derives from diseases associated to SNPs, where the correction of the genomic site through HDR requires nuclease cleavage in close proximity to the target nucleotide³⁶⁴. Similarly, the correction of highly homologous genes involved in diseases (i.e. SMN, Ig and TCR genes) significantly limits the choice of sgRNAs or requires high levels of specificity to discriminate between the target site and its very similar homologous copies. Thus, where off-target cleavages cannot be avoided by a careful choice of the guide RNA sequence due to experimental limitations, Cas9-mediated editing would certainly benefit from those strategies that allow to increase the nuclease on-/off-target ratio. Several methods to reduce Cas9 off-target activity have been designed in the past years and a major step forward towards the solution of this issue was the development of high-fidelity Cas9 variants. Two groups used rational structure-guided protein engineering to design mutants with increased specificity^{347,348}, however we believe that the unbiased screening approach we employed in the present study allows the identification of non-trivial substitutions leading to a further increase in specificity, as demonstrated by the isolation of evoCas9, as well as other alternative variants we did not characterize in detail yet. In addition, once built, our yeast platform can be easily adapted with minimal modifications to screen for improved variants of other RNA-guided nucleases as well as for the identification of additional mutations reducing off-target activity in other SpCas9 domains.

When compared side-by-side with eSpCas9(1.1) and SpCas9-HF1 using an EGFP knockout reporter assay, evoCas9 demonstrated superior performance in off-target discrimination, with the ability to almost completely avoid the cleavage of a sequence containing a single mismatch in a very PAM-distal position, a feature completely absent from previously published mutants. Accordingly, when the same assay was repeated prolonging evoCas9 expression by lentiviral transduction, the differences with the wild-type protein and SpCas9-HF1 were even more pronounced, although the control of the same PAM-distal mismatched target was only partial. Similarly, when we tested evoCas9 off-target activity against endogenous loci we could detect cleavage only for sites differing from the intended target by one nucleotide (FANCF site 2 and CCR5) or relative to highly repetitive target sequences (VEGFA site 2 and site 3). Overall, these results indicate that careful sgRNA selection, when possible, is crucial to avoid residual levels of unspecific cleavage and the exploitation of widely available online sgRNA design tools can help in predicting possible obvious matches in the target genome. Importantly, in our experimental conditions evoCas9 was generally characterized by a further 2 to 4-fold reduction in off-target activity if compared to SpCas9-HF1 (**Fig. 20** and **Fig. 24**), that already reduced unwanted cleavages to background levels for many, but not all, of the tested sites³⁴⁷. In particular, GUIDE-seq analysis of the VEGFA site 2 locus off-target sites captured only 10 sites for evoCas9, while identifying around 600 loci for the wild-type nuclease: previously published results reported ~20 sites for SpCas9-HF1 together with only ~150 sites for wild-type Cas9³⁴⁷, most likely indicating lower sequencing depth that may have led to an underestimation of SpCas9-HF1 off-target activity.

The characterization of evoCas9 on-target activity revealed a partial loss of editing for some of the endogenous tested sites that was particularly pronounced for the ZSCAN2 locus. This effect could indicate a trade-off between specificity and cleavage activity, where the reduction of the binding and cleavage of off-target sites is paralleled to a lesser extent by a similar effect on the on-target site, and has been observed also in previously published reports³⁴⁷. The combination of evoCas9 with optimized guide RNAs, which we demonstrated to be compatible with our variant, may help in reverting this loss of activity given their general ability to improve SpCas9 functionality³⁶¹. Our yeast screen allowed the identification of several substitutions that may be further

combined to generate alternative high-fidelity variants with a different balance in the trade-off between specificity and activity. In addition, further screening using our yeast platform could lead to the identification of other mutations not yet discovered in the present study. Alternatively, the extension of the duration of SpCas9 expression through stable transduction of target cells may increase the overall editing efficiency. As we have demonstrated, evoCas9 is indeed able to avoid cleavage of at least some off-target sites even after prolonged expression into cells (**Fig. 28**).

Additional investigations will be necessary to understand the mechanism behind evoCas9 increased specificity. Preliminary observations obtained using an evoCas9-based transcriptional activator revealed that our variant bound DNA less efficiently than the wild-type protein. This suggests that the modifications introduced in evoCas9 reduce the strength of the non-sequence specific interactions between the nuclease and the target DNA backbone. Binding and cleavage will be thus more dependent on the perfect base pairing of the spacer RNA sequence with the target DNA. Further biochemical characterization will clarify the kinetic details of evoCas9 interaction with the target DNA and of the cleavage reaction. The solution of the crystal structure of evoCas9 in complex with a mismatched target DNA or *in silico* modelling approaches starting from available structural data will provide important insights on the effects of the substitutions introduced in the variant on off-target recognition. In addition, valuable information could be gathered to explain the particular behaviour observed when evoCas9 was tested with sgRNA longer than 20 nucleotides or containing a mismatched guanine at the 5'-end of the spacer sequence (position 20), leading to a complete loss of cleavage activity.

Finally, it is possible to speculate that a further improvement of targeting specificity could be obtained by the combination of evoCas9 with some of the previously proposed methods to reduce off-target activity, such as paired Cas9 nickases³²⁵ or RNP delivery. However, there is no absolute compatibility, as we demonstrated that tru-gRNAs³²² cannot be combined with evoCas9 without a significant drop in cleavage efficiency.

Permanent expression of Cas9 into target cells represents an issue that in some cases even SpCas9 mutants with highly increased fidelity, such as evoCas9, may not be able to handle. Different methods, such as RNP delivery or RNA transfection, have been

proposed as a way to reduce Cas9 permanence into cells. However, these delivery methods are highly inefficient for *in vivo* treatments. Our self-limiting circuit strategy, lentiSLiCES, exploits the efficiency of viral based delivery and simultaneously limits the intracellular amount of SpCas9 post-transduction and viral integration. By limiting in time and abundance Cas9 expression, SLiCES avoids the accumulation of off-target cleavages that instead are observed with the use of conventional Cas9 delivery approaches. The use of integrating lentiviral vectors coupled to our self-targeting strategy may rise safety concerns associated with the potential introduction of unwanted chromosomal translocations between multiple copies of integrated vectors after self-editing. Nevertheless, fine tuning of the amount of vector used should limit the amount of integration to a single event per cell. In addition, the quick kinetics of Cas9 downregulation (see **Fig. 36e**), may suggest that self-editing of the Cas9 transgene occurs before integration, thus limiting the damage at the level of the cellular genomic DNA. Furthermore, to improve the safety of the SLiCES strategy, Integrase Defective Lentiviral Vectors (IDLV)³⁶⁵ could be used to maintain the viral-based efficiency in cellular delivery, while avoiding any possible issue related to lentiviral integration.

A variety of Cas9 applications, such as the regulation of gene expression obtained by the combination with transcriptional activation domains^{262,265,323} might be significantly improved through their adaptation to lentiSLiCES. In fact, these approaches as well as the refined modulation of gene expression obtained with a genetic kill-switch circuit^{366,367} could be potentiated by a tunable self-limiting approach to restrict in time Cas9-mediated induction of the targeted cellular promoters. Finally, SLiCES may significantly improve some recently developed Cas9 genome engineering procedures that are susceptible to continuous nuclease activity. For instance, current techniques to efficiently substitute genomic sequences use Cas9 to increase the rate of homology-directed repair; nevertheless, these techniques are often limited by the continuous re-cleavage of the newly substituted genomic sequence by Cas9³⁶⁴. Even though modification of the exogenous donor template by introducing silent mutations in the targeted sequence or at the PAM level may help in avoiding re-editing by Cas9, it is not always an applicable strategy. This may occur where the sequence of nucleotides surrounding the edited base or region has a biological significance and cannot be thus

modified (e.g. non coding sequences, binding sites). In all these cases the nuclease inactivation strategy presented in this thesis work represents a valid alternative.

Similar approaches aimed at controlling Cas9 activity have been recently developed by exploiting various inducible systems³⁴¹. Nevertheless, the methodologies reported so far suffer from a number of limitations going from decreased editing activity generated by nuclease splitting³⁶⁸ or chemical modification³⁴² to background activity³⁴⁴ or extended time of required induction³⁴³.

lentiSLiCES offers an efficient delivery tool that after editing creates a nuclease-free cellular environment, greatly improving the safety margins of the CRISPR technology. Notably, at the systemic level, a pre-existing immunity against SpCas9 was recently proved in mice³⁶⁹ suggesting that *in vivo* approaches will be severely limited by the immune response against the nuclease; thus the transient nature of the SLiCES system may have a significant impact for CRISPR clinical use.

The development of evoCas9, characterized by a highly increased fidelity, aims to address pressing safety issues associated with the use of Cas9 into the clinic³⁷⁰, in order to contribute to fill the gap between bench and bedside for this promising technology. The increased specificity of evoCas9 for the first time ever could allow the design of therapeutic strategies based on the editing either by NHEJ or HDR of highly homologous sequences, differing only for one or few bases, for example when only one pathogenic allele must be targeted.

Overall, the “hit and go” nature of SLiCES and its adaptability to new emerging Cas9 techniques, combined with the implementation of its viral delivery, allows more controllable genome editing procedures with limited unwanted off-target activity. On the other hand, the identification of evoCas9, that retains near-wild type levels of on-target cleavage while reducing off-target activity to levels below those obtained with the best performing variants reported so far, represents a step towards the generation of a CRISPR toolbox more suitable for therapeutic applications as well as for basic research. Furthermore, it is possible to envision the combination of lentiSLiCES with evoCas9 to obtain an all-in-one safe and error-free genome editing platform for more reliable *in vivo* approaches.

Contribution to the experiments

All the experiments not specifically mentioned below were performed by myself.

Identification of high-fidelity SpCas9 variants:

The experiments involving the generation of the yeast reporter strains for the evaluation of Cas9 on- and off-target activity, the screening for the selection of high-fidelity SpCas9 variants in yeast and their validation in the EGFP-knockout mammalian system were performed together with Michele Olivieri, during his Master thesis and some months after his graduation.

Library preparation for GUIDE-seq off-target analysis was performed by Francesca Lorenzin.

Bioinformatics analyses for the GUIDE-seq experiments were performed by Davide Prandi, while the analyses related to targeted deep-sequencing experiments for off-target activity evaluation were carried out by Alessandro Romanel.

Self-limiting Cas9 circuits:

The experiments involving the transient transfection of plasmid DNA for the set up and the initial evaluation of SLiCES behaviour were performed by Gianluca Petris, who also helped to conduct the TIDE analyses to measure lentiSLiCES specificity against endogenous loci.

Acknowledgments

These have been five long years, with all their ups and downs. These have also been five intense years and the people I met and to whom I have to be grateful are quite a number. As a matter of fact, I now realize how lucky I was to end up in such a welcoming and stimulating environment. I believe that this is the spirit of CIBIO, which I experienced on my own skin: creating a unique setting to foster scientific talent and personal growth, without the burden of too strict hierarchies or excessive formality. Well, this is how I imagined doing Science would be: ideas count, not the position of the person who proposes them. Mixing people means sharing thoughts and knowledge...who knows then what may come out of the pot! In that CIBIO is special – and will remain my personal benchmark reference for all the future places I will work in.

When I started my PhD I had a couple of rough years and this confirmed what I already knew: doing Science is difficult, sometimes even frustrating. However, failures help to make a good scientist, I think. Troubleshooting, plans B (up to Z) together with a good dose of perseverance and resilience are what it takes to reach your goal. In a way or another you have to learn that. The silver lining here is that the satisfaction of finally telling a good story, that feeling when you realise that you managed to crack a tiny slit in that black box which is Life to have a quick peep inside, washes away instantly any weariness. And that is, more or less, my story.

Many people accompanied me during my PhD journey. Most of all I have to thank my mentor Anna Cereseto for allowing me to join her lab once more. Without her support nothing of this would have been possible and to her goes my gratitude for believing in me as a scientist and for nurturing my growth during these years. If it was good fun to work here I have to thank my former and present lab mates: Ashwanth known as Francis, who welcomed me in CIBIO; Tiziana, Greta and Ruben. Then Gianluca, bursting with new ideas every day; Steve, with his fantastic English humour; Claudia, always cheerful and calm; Giulia, who will take over the responsibility as the Cereseto's Lab PhD student (good luck!). Michele and Giordano, that from students became good friends, it was really a pleasure to work with you. I am sure that both will do great things. Thank you all for putting up with me for 12 hours a day (almost) in the past

years. Thank you for creating a stimulating and productive environment, I couldn't have asked for better companions.

A special thank goes to all my friends in CIBIO, the old ones and always present, who shared with me the joys and the dark moments of our PhD life since the very beginning, the gone ones and the new. Thank you for cheering up everyday life and for your support in the more difficult moments. I am really lucky to have met you all "Amici della Croazia". Willing or not, you were main players of my life.

I want to thank Paolo Macchi, the first director of our PhD Program, for creating this unique environment, it was really a pleasure to be a PhD student here. Together with him, I want also to express my gratitude to Betty Balduin as she was always supportive and present for helping me going through all the deadlines and mandatory activities.

I would like also to acknowledge Luca Andreotti, as well as the rest of the lab management team, for many of the experiments I have performed in these five years went smoothly thanks to their tireless work to keep CIBIO fully operating.

I am sure to have forgotten someone, and anyway it would probably take pages and pages to properly acknowledge all the people who deserve to be thanked, so I will simply finish by saying that I am grateful to have spent these five years in CIBIO and CIBIO is made of people, first of all.

These, however, would be terribly incomplete acknowledgements. Starting a PhD is not something you do out of the blue. Facing failures and disappointments is sometimes very difficult. Finding new energies and motivation to enter the lab can be challenging sometimes. This is why I want to thank my (extended) family, especially my parents, and my wonderful girlfriend Veronica. Your support and your encouragement kept me going in the most difficult moments during these years. You gave me the motivation to always strive for improvement. As did my early morning conversations with Vittorio, unmissable and inspiring. To be fair, I must also thank Veronica for her patience and understanding: I bet it is not always easy to cope with my working habits.

I clearly see how lucky I am to have you all and I will do whatever it takes to deserve your respect and your continuous support.

Bibliography

1. Larson, G. *et al.* Rethinking dog domestication by integrating genetics, archeology, and biogeography. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 8878–8883 (2012).
2. Avery, O. T., Macleod, C. M. & McCarty, M. Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types: Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated From Pneumococcus Type III. *J. Exp. Med.* **79**, 137–158 (1944).
3. Cohen, S. N., Chang, A. C., Boyer, H. W. & Helling, R. B. Construction of biologically functional bacterial plasmids in vitro. *Proceedings of the National Academy of Sciences* **70**, 3240–3244 (1973).
4. Jaenisch, R. & Mintz, B. Simian Virus 40 DNA Sequences in DNA of Healthy Adult Mice Derived from Preimplantation Blastocysts Injected with Viral DNA. *Proceedings of the National Academy of Sciences* **71**, 1250–1254 (1974).
5. Gibson, D. G. *et al.* Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329**, 52–56 (2010).
6. Hutchison, C. A. *et al.* Design and synthesis of a minimal bacterial genome. *Science* **351**, aad6253 (2016).
7. Ronald, P. Plant Genetics, Sustainable Agriculture and Global Food Security. *Genetics* **188**, 11–20 (2011).
8. Friedmann, T. & Roblin, R. Gene Therapy for Human Genetic Disease? *Science* **175**, 949–955 (1972).
9. Naldini, L. Gene therapy returns to centre stage. *Nature* **526**, 351–360 (2015).
10. Nayerossadat, N., Ali, P. & Maedeh, T. Viral and nonviral delivery systems for gene delivery. *Adv Biomed Res* **1**, 27–11 (2012).
11. Thomas, C. E., Ehrhardt, A. & Kay, M. A. Progress and problems with the use of viral vectors for gene therapy. *Nat. Rev. Genet.* **4**, 346–358 (2003).
12. Kotterman, M. A., Chalberg, T. W. & Schaffer, D. V. Viral Vectors for Gene Therapy: Translational and Clinical Outlook. *Annu. Rev. Biomed. Eng.* **17**, 63–89 (2015).
13. O'Connor, T. P. & Crystal, R. G. Genetic medicines: treatment strategies for hereditary disorders. *Nat. Rev. Genet.* **7**, 261–276 (2006).
14. Ylä-Herttua, S. Endgame: Glybera Finally Recommended for Approval as the First Gene Therapy Drug in the European Union. *Molecular Therapy* **20**, 1831–1832 (2012).
15. Gaudet, D. *et al.* Efficacy and long-term safety of alipogene tiparvovec (AAV1-LPLS447X) gene therapy for lipoprotein lipase deficiency: an open-label trial. *Gene Ther* **20**, 361–369 (2013).
16. Schimmer, J. & Breazzano, S. Investor Outlook: Rising from the Ashes; GSK's European Approval of Strimvelis for ADA-SCID. *Human Gene Therapy Clinical Development* **27**, 57–61 (2016).
17. Aiuti, A. *et al.* Gene therapy for immunodeficiency due to adenosine deaminase deficiency. *N Engl J Med* **360**, 447–458 (2009).
18. Baum, C., Kustikova, O., Modlich, U., Li, Z. X. & Fehse, B. Mutagenesis and oncogenesis by chromosomal insertion of gene transfer vectors. *Human Gene Therapy* **17**, 253–263 (2006).
19. Cavazzana-Calvo, M. *et al.* Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease. *Science* **288**, 669–672 (2000).
20. Hacein-Bey-Abina, S. LMO2-Associated Clonal T Cell Proliferation in Two Patients after Gene Therapy for SCID-X1. *Science* **302**, 415–419 (2003).
21. Kohn, D. B., Sadelain, M. & Glorioso, J. C. Occurrence of leukaemia following gene therapy of X-linked SCID. *Nat Rev Cancer* **3**, 477–488 (2003).

22. Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J. & Stahl, F. W. The double-strand-break repair model for recombination. *Cell* **33**, 25–35 (1983).
23. Rothstein, R. J. One-step gene disruption in yeast. *Meth. Enzymol.* **101**, 202–211 (1983).
24. Capecchi, M. R. Generating mice with targeted mutations. *Nat Med* **7**, 1086–1090 (2001).
25. Capecchi, M. R. The new mouse genetics: altering the genome by gene targeting. *Trends Genet.* **5**, 70–76 (1989).
26. Capecchi, M. R. Altering the genome by homologous recombination. *Science* (1989).
27. Thomas, K. R. & Capecchi, M. R. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* **51**, 503–512 (1987).
28. Doetschman, T. *et al.* Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. *Nature* **330**, 576–578 (1987).
29. Folger, K. R., Wong, E. A., Wahl, G. & Capecchi, M. R. Patterns of integration of DNA microinjected into cultured mammalian cells: evidence for homologous recombination between injected plasmid DNA molecules. *Molecular and Cellular Biology* **2**, 1372–1387 (1982).
30. Ayares, D., Chekuri, L., Song, K. Y. & Kucherlapati, R. Sequence homology requirements for intermolecular recombination in mammalian cells. *Proceedings of the National Academy of Sciences* **83**, 5199–5203 (1986).
31. Thomas, K. R., Folger, K. R. & Capecchi, M. R. High frequency targeting of genes to specific sites in the mammalian genome. *Cell* **44**, 419–428 (1986).
32. Wong, E. A. & Capecchi, M. R. Homologous recombination between coinjected DNA sequences peaks in early to mid-S phase. *Molecular and Cellular Biology* **7**, 2294–2295 (1987).
33. Takata, M. *et al.* Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. *The EMBO Journal* **17**, 5497–5508 (1998).
34. Folger, K. R., Thomas, K. & Capecchi, M. R. Nonreciprocal exchanges of information between DNA duplexes coinjected into mammalian cell nuclei. *Molecular and Cellular Biology* **5**, 59–69 (1985).
35. Valancius, V. & Smithies, O. Testing an ‘in-out’ targeting procedure for making subtle genomic modifications in mouse embryonic stem cells. *Molecular and Cellular Biology* **11**, 1402–1408 (1991).
36. Moore, R. C. *et al.* Double replacement gene targeting for the production of a series of mouse strains with different prion protein gene alterations. *Biotechnology (N.Y.)* **13**, 999–1004 (1995).
37. Kos, C. H. Cre/loxP system for generating tissue-specific knockout mouse models. *Nutr. Rev.* **62**, 243–246 (2004).
38. Kim, H. & Kim, J.-S. A guide to genome engineering with programmable nucleases. *Nature Publishing Group* **15**, 321–334 (2014).
39. Rouet, P., Smih, F. & Jasin, M. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Molecular and Cellular Biology* **14**, 8096–8106 (1994).
40. Smih, F., Rouet, P., Romanienko, P. J. & Jasin, M. Double-strand breaks at the target locus stimulate gene targeting in embryonic stem cells. *Nucleic Acids Research* **23**, 5012–5019 (1995).
41. Rouet, P., Smih, F. & Jasin, M. Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proceedings of the National Academy of Sciences* **91**, 6064–6068 (1994).
42. Urnov, F. D. *et al.* Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* **435**, 646–651 (2005).
43. Maeder, M. L. & Gersbach, C. A. Genome Editing Technologies for Gene and Cell

- Therapy. *Molecular Therapy* 1–17 (2016). doi:10.1038/mt.2016.10
44. Bedell, V. M. *et al.* In vivo genome editing using a high-efficiency TALEN system. *Nature* 1–7 (2012). doi:10.1038/nature11537
 45. Chen, F. *et al.* High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. *Nat. Methods* **8**, 753–755 (2011).
 46. Ding, Q. *et al.* A TALEN Genome-Editing System for Generating Human Stem Cell-Based Disease Models. *Stem Cell* 1–14 (2012). doi:10.1016/j.stem.2012.11.011
 47. Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. & Corn, J. E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat. Biotechnol.* **34**, 1–7 (2016).
 48. Genovese, P. *et al.* Targeted genome editing in human repopulating haematopoietic stem cells. *Nature* **510**, 235–240 (2014).
 49. Händel, E.-M. *et al.* Versatile and Efficient Genome Editing in Human Cells by Combining Zinc-Finger Nucleases With Adeno-Associated Viral Vectors. *Human Gene Therapy* **23**, 321–329 (2012).
 50. Coluccio, A. *et al.* Targeted Gene Addition in Human Epithelial Stem Cells by Zinc-finger Nuclease-mediated Homologous Recombination. *Molecular Therapy* **21**, 1695–1704 (2013).
 51. Hirsch, M. L., Green, L., Porteus, M. H. & Samulski, R. J. Self-complementary AAV mediates gene targeting and enhances endonuclease delivery for double-strand break repair. *Gene Ther* **17**, 1175–1180 (2010).
 52. Holkers, M. *et al.* Adenoviral vector DNA for accurate genome editing with engineered nucleases. *Nat. Methods* 1–11 (2014). doi:10.1038/nmeth.3075
 53. Joglekar, A. V. *et al.* Integrase-defective Lentiviral Vectors as a Delivery Platform for Targeted Modification of Adenosine Deaminase Locus. *Molecular Therapy* **21**, 1705–1717 (2013).
 54. Moehle, E. A. *et al.* Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. *Proceedings of the National Academy of Sciences* **104**, 3055–3060 (2007).
 55. Lombardo, A. *et al.* Site-specific integration and tailoring of cassette design for sustainable gene transfer. *Nat. Methods* **8**, 861–869 (2011).
 56. Li, H. *et al.* In vivo genome editing restores haemostasis in a mouse model of haemophilia. *Nature* **475**, 217–221 (2011).
 57. Lieber, M. R. The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End-Joining Pathway. *Annu. Rev. Biochem.* **79**, 181–211 (2010).
 58. Ciccia, A. & Elledge, S. J. The DNA Damage Response: Making It Safe to Play with Knives. *Molecular Cell* **40**, 179–204 (2010).
 59. Chapman, J. R., Taylor, M. R. G. & Boulton, S. J. Playing the end game: DNA double-strand break repair pathway choice. *Molecular Cell* **47**, 497–510 (2012).
 60. Lee, H. J., Kim, E. & Kim, J. S. Targeted chromosomal deletions in human cells using zinc finger nucleases. *Genome Research* **20**, 81–89 (2010).
 61. Xiao, A. *et al.* Chromosomal deletions and inversions mediated by TALENs and CRISPR/Cas in zebrafish. *Nucleic Acids Research* **41**, e141–e141 (2013).
 62. Canver, M. C. *et al.* Characterization of genomic deletion efficiency mediated by clustered regularly interspaced palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells. *J. Biol. Chem.* **289**, 21312–21324 (2014).
 63. Sollu, C. *et al.* Autonomous zinc-finger nuclease pairs for targeted chromosomal deletion. *Nucleic Acids Research* **38**, 8269–8276 (2010).
 64. Ousterout, D. G. *et al.* Multiplex CRISPR/Cas9-based genome editing for correction of dystrophin mutations that cause Duchenne muscular dystrophy. *Nat Comms* **6**, 6244 (2015).
 65. Nelson, C. E. *et al.* In vivo genome editing improves muscle function in a mouse model

- of Duchenne muscular dystrophy. *Science* **351**, 403–407 (2016).
66. Aronin, N. & DiFiglia, M. Huntingtin-lowering strategies in Huntington's disease: Antisense oligonucleotides, small RNAs, and gene editing. *Mov Disord.* **29**, 1455–1461 (2014).
 67. 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 (2014).
 68. Perez, E. E. *et al.* Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nat. Biotechnol.* **26**, 808–816 (2008).
 69. Holt, N. *et al.* Human hematopoietic stem/progenitor cells modified by zinc-finger nucleases targeted to CCR5 control HIV-1 in vivo. *Nat. Biotechnol.* **28**, 839–847 (2010).
 70. de Silva, E. & Stumpf, M. P. H. HIV and the CCR5-Delta32 resistance allele. *FEMS Microbiol. Lett.* **241**, 1–12 (2004).
 71. Chevalier, B. S. & Stoddard, B. L. Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Research* **29**, 3757–3774 (2001).
 72. Hafez, M., Hausner, G. & Bonen, L. Homing endonucleases: DNA scissors on a mission. *Genome* **55**, 553–569 (2012).
 73. Belfort, M. & Roberts, R. J. Homing endonucleases: keeping the house in order. *Nucleic Acids Research* **25**, 3379–3388 (1997).
 74. Edgell, D. R., Chalamcharla, V. R. & Belfort, M. Learning to live together: mutualism between self-splicing introns and their hosts. *BMC Biol* **9**, 22 (2011).
 75. Silva, G. *et al.* Meganucleases and other tools for targeted genome engineering: perspectives and challenges for gene therapy. *Curr Gene Ther* **11**, 11–27 (2011).
 76. Prieto, J., Molina, R. & Montoya, G. Molecular scissors for in situ cellular repair. *Crit. Rev. Biochem. Mol. Biol.* **47**, 207–221 (2012).
 77. Chen, Z., Wen, F., Sun, N. & Zhao, H. Directed evolution of homing endonuclease I-SceI with altered sequence specificity. *Protein Eng. Des. Sel.* **22**, 249–256 (2009).
 78. Doyon, J. B., Pattanayak, V., Meyer, C. B. & Liu, D. R. Directed Evolution and Substrate Specificity Profile of Homing Endonuclease I-SceI. *J. Am. Chem. Soc.* **128**, 2477–2484 (2006).
 79. Rosen, L. E. *et al.* Homing endonuclease I-CreI derivatives with novel DNA target specificities. *Nucleic Acids Research* **34**, 4791–4800 (2006).
 80. Seligman, L. M. *et al.* Mutations altering the cleavage specificity of a homing endonuclease. *Nucleic Acids Research* **30**, 3870–3879 (2002).
 81. Sussman, D. *et al.* Isolation and Characterization of New Homing Endonuclease Specificities at Individual Target Site Positions. *Journal of Molecular Biology* **342**, 31–41 (2004).
 82. Epinat, J. C. A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Research* **31**, 2952–2962 (2003).
 83. Chevalier, B. S. *et al.* Design, activity, and structure of a highly specific artificial endonuclease. *Molecular Cell* **10**, 895–905 (2002).
 84. Grizot, S. *et al.* Generation of redesigned homing endonucleases comprising DNA-binding domains derived from two different scaffolds. *Nucleic Acids Research* **38**, 2006–2018 (2010).
 85. Arnould, S. *et al.* Engineering of Large Numbers of Highly Specific Homing Endonucleases that Induce Recombination on Novel DNA Targets. *Journal of Molecular Biology* **355**, 443–458 (2006).
 86. McConnell Smith, A. *et al.* Generation of a nicking enzyme that stimulates site-specific gene conversion from the I-Anil LAGLIDADG homing endonuclease. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 5099–5104 (2009).
 87. Niu, Y., Tenney, K., Li, H. & Gimble, F. S. Engineering Variants of the I-SceI Homing Endonuclease with Strand-specific and Site-specific DNA-nicking Activity. *Journal of*

- Molecular Biology* **382**, 188–202 (2008).
88. Puchta, H., Dujon, B. & Hohn, B. Two different but related mechanisms are used in plants for the repair of genomic double-strand breaks by homologous recombination. *Proceedings of the National Academy of Sciences* **93**, 5055–5060 (1996).
 89. Thermes, V. *et al.* I-SceI meganuclease mediates highly efficient transgenesis in fish. *Mech. Dev.* **118**, 91–98 (2002).
 90. Glover, L., McCulloch, R. & Horn, D. Sequence homology and microhomology dominate chromosomal double-strand break repair in African trypanosomes. *Nucleic Acids Research* **36**, 2608–2618 (2008).
 91. Maggert, K. A., Gong, W. J. & Golic, K. G. Methods for homologous recombination in *Drosophila*. *Methods Mol. Biol.* **420**, 155–174 (2008).
 92. Windbichler, N. *et al.* Homing endonuclease mediated gene targeting in *Anopheles gambiae* cells and embryos. *Nucleic Acids Research* **35**, 5922–5933 (2007).
 93. Gouble, A. *et al.* Efficient in toto targeted recombination in mouse liver by meganuclease-induced double-strand break. *J. Gene Med.* **8**, 616–622 (2006).
 94. Redondo, P. *et al.* Molecular basis of xeroderma pigmentosum group C DNA recognition by engineered meganucleases. *Nature* **456**, 107–111 (2008).
 95. Grizot, S. *et al.* Efficient targeting of a SCID gene by an engineered single-chain homing endonuclease. *Nucleic Acids Research* **37**, 5405–5419 (2009).
 96. Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S. & Gregory, P. D. Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.* **11**, 636–646 (2010).
 97. Kim, Y. G., Cha, J. & Chandrasegaran, S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences* **93**, 1156–1160 (1996).
 98. Vanamee, É. S., Santagata, S. & Aggarwal, A. K. FokI requires two specific DNA sites for cleavage. *Journal of Molecular Biology* **309**, 69–78 (2001).
 99. Szczepek, M. *et al.* Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nat. Biotechnol.* **25**, 786–793 (2007).
 100. Miller, J. C. *et al.* An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat. Biotechnol.* **25**, 778–785 (2007).
 101. Händel, E.-M., Alwin, S. & Cathomen, T. Expanding or restricting the target site repertoire of zinc-finger nucleases: the inter-domain linker as a major determinant of target site selectivity. *Mol. Ther.* **17**, 104–111 (2009).
 102. Wolfe, S. A., Nekludova, L. & Pabo, C. O. DNA recognition by Cys2His2 zinc finger proteins. *Annual review of biophysics ...* **29**, 183–212 (2000).
 103. Pavletich, N. P. & Pabo, C. O. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* **252**, 809–817 (1991).
 104. Segal, D. J. *et al.* Evaluation of a Modular Strategy for the Construction of Novel Polydactyl Zinc Finger DNA-Binding Proteins †. *Biochemistry* **42**, 2137–2148 (2003).
 105. Choo, Y. & Klug, A. Toward a code for the interactions of zinc fingers with DNA: selection of randomized fingers displayed on phage. *Proceedings of the National Academy of Sciences* **91**, 11163–11167 (1994).
 106. Rebar, E. J. & Pabo, C. O. Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. *Science* **263**, 671–673 (1994).
 107. Fairall, L., Schwabe, J. W., Chapman, L., Finch, J. T. & Rhodes, D. The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition. *Nature* **366**, 483–487 (1993).
 108. Wolfe, S. A., Grant, R. A., Elrod-Erickson, M. & Pabo, C. O. Beyond the 'recognition code': structures of two Cys2His2 zinc finger/TATA box complexes. *Structure* (2001).
 109. Pavletich, N. P. & Pabo, C. O. Crystal structure of a five-finger GLI-DNA complex: new perspectives on zinc fingers. *Science* **261**, 1701–1707 (1993).
 110. Foley, J. E. *et al.* Rapid Mutation of Endogenous Zebrafish Genes Using Zinc Finger

- Nucleases Made by Oligomerized Pool ENgineering (OPEN). *PLoS ONE* **4**, e4348–13 (2009).
111. Meng, X., Noyes, M. B., Zhu, L. J., Lawson, N. D. & Wolfe, S. A. Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nat. Biotechnol.* **26**, 695–701 (2008).
 112. Sander, J. D. *et al.* Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nat. Methods* **8**, 67–69 (2010).
 113. Moore, M., Klug, A. & Choo, Y. Improved DNA binding specificity from polyzinc finger peptides by using strings of two-finger units. *Proceedings of the National Academy of Sciences* **98**, 1437–1441 (2001).
 114. Bibikova, M., Golic, M., Golic, K. G. & Carroll, D. Targeted chromosomal cleavage and mutagenesis in *Drosophila* using zinc-finger nucleases. *Genetics* **161**, 1169–1175 (2002).
 115. Doyon, Y. *et al.* Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nat. Biotechnol.* **26**, 702–708 (2008).
 116. Geurts, A. M. *et al.* Knockout Rats via Embryo Microinjection of Zinc-Finger Nucleases. *Science* **325**, 433–433 (2009).
 117. Lloyd, A., Plaisier, C. L., Carroll, D. & Drews, G. N. Targeted mutagenesis using zinc-finger nucleases in *Arabidopsis*. *Proceedings of the National Academy of Sciences* **102**, 2232–2237 (2005).
 118. Joung, J. K. & Sander, J. D. TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol* 1–7 (2012). doi:10.1038/nrm3486
 119. Santiago, Y. *et al.* Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 5809–5814 (2008).
 120. Liu, P.-Q. *et al.* Generation of a triple-gene knockout mammalian cell line using engineered zinc-finger nucleases. *Biotechnol. Bioeng.* n/a–n/a (2010). doi:10.1002/bit.22654
 121. Lombardo, A. *et al.* Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nat. Biotechnol.* **25**, 1298–1306 (2007).
 122. Hockemeyer, D. *et al.* Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat. Biotechnol.* **27**, 851–857 (2009).
 123. Zou, J. *et al.* Gene Targeting of a Disease-Related Gene in Human Induced Pluripotent Stem and Embryonic Stem Cells. *Stem Cell* **5**, 97–110 (2009).
 124. Chou, S.-T., Leng, Q. & Mixson, A. J. Zinc Finger Nucleases: Tailor-made for Gene Therapy. *Drugs Future* **37**, 183–196 (2012).
 125. Sebastiano, V. *et al.* In Situ Genetic Correction of the Sickle Cell Anemia Mutation in Human Induced Pluripotent Stem Cells Using Engineered Zinc Finger Nucleases. *STEM CELLS* **29**, 1717–1726 (2011).
 126. Yusa, K. *et al.* Targeted gene correction of α 1-antitrypsin deficiency in induced pluripotent stem cells. *Nature* **478**, 391–394 (2011).
 127. Soldner, F. *et al.* Generation of Isogenic Pluripotent Stem Cells Differing Exclusively at Two Early Onset Parkinson Point Mutations. *Cell* **146**, 318–331 (2011).
 128. Dudley, M. E., Wunderlich, J. R. & Yang, J. C. A phase I study of nonmyeloablative chemotherapy and adoptive transfer of autologous tumor antigen-specific T lymphocytes in patients with metastatic *Journal of ...* (2002). doi:10.1097/01.CJI.0000016820.36510.89
 129. Hütter, G. *et al.* Long-term control of HIV by CCR5 Delta32/Delta32 stem-cell transplantation. *N Engl J Med* **360**, 692–698 (2009).
 130. Sun, N. & Zhao, H. Transcription activator-like effector nucleases (TALENs): A highly efficient and versatile tool for genome editing. *Biotechnol. Bioeng.* **110**, 1811–1821 (2013).
 131. Cade, L. *et al.* Highly efficient generation of heritable zebrafish gene mutations using homo- and heterodimeric TALENs. *Nucleic Acids Research* **40**, 8001–8010 (2012).
 132. Boch, J. & Bonas, U. Xanthomonas AvrBs3 Family-Type III Effectors: Discovery and

- Function. *Annu. Rev. Phytopathol.* **48**, 419–436 (2010).
133. Boch, J. *et al.* Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. *Science* **326**, 1509–1512 (2009).
 134. Sun, N., Liang, J., Abil, Z. & Zhao, H. Optimized TAL effector nucleases (TALENs) for use in treatment of sickle cell disease. *Mol. BioSyst.* **8**, 1255–9 (2012).
 135. Miller, J. C. *et al.* A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.* **29**, 143–148 (2010).
 136. Li, T. *et al.* TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Research* **39**, 359–372 (2011).
 137. Christian, M. *et al.* Targeting DNA Double-Strand Breaks with TAL Effector Nucleases. *Genetics* **186**, 757–761 (2010).
 138. Mussolino, C. *et al.* A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Research* **39**, 9283–9293 (2011).
 139. Christian, M. L. *et al.* Targeting G with TAL Effectors: A Comparison of Activities of TALENs Constructed with NN and NK Repeat Variable Di-Residues. *PLoS ONE* **7**, e45383–9 (2012).
 140. Gao, H., Wu, X., Chai, J. & Han, Z. Crystal structure of a TALE protein reveals an extended N-terminal DNA binding region. *Cell Research* **22**, 1716–1720 (2012).
 141. Moscou, M. J. & Bogdanove, A. J. A Simple Cipher Governs DNA Recognition by TAL Effectors. *Science* **326**, 1501–1501 (2009).
 142. Mak, A. N. S., Bradley, P., Cernadas, R. A., Bogdanove, A. J. & Stoddard, B. L. The Crystal Structure of TAL Effector PthXo1 Bound to Its DNA Target. *Science* **335**, 716–719 (2012).
 143. Deng, D. *et al.* Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science* **335**, 720–723 (2012).
 144. Morbitzer, R., Römer, P. & Boch, J. Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors. in (2010). doi:10.1073/pnas.1013133107/-/DCSupplemental
 145. Le Cong, Zhou, R., Kuo, Y.-C., Cunniff, M. & Zhang, F. Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. *Nat Comms* **3**, 968–6 (1AD).
 146. Bultmann, S. *et al.* Targeted transcriptional activation of silent oct4 pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers. *Nucleic Acids Research* **40**, 5368–5377 (2012).
 147. Deng, D. *et al.* Recognition of methylated DNA by TAL effectors. *Cell Research* **22**, 1502–1504 (2012).
 148. Valton, J. *et al.* Overcoming transcription activator-like effector (TALE) DNA binding domain sensitivity to cytosine methylation. *J. Biol. Chem.* **287**, 38427–38432 (2012).
 149. Huang, P. *et al.* Heritable gene targeting in zebrafish using customized TALENs. *Nat. Biotechnol.* **29**, 699–700 (2011).
 150. Reyon, D., Khayter, C., Regan, M. R. & Joung, J. K. Engineering designer transcription activator-like effector nucleases (TALENs). *Current protocols in ...* (2012). doi:10.1002/0471142727.mb1215s100
 151. Sander, J. D. *et al.* Targeted gene disruption in somatic zebrafish cells using engineered TALENs. *Nat. Biotechnol.* **29**, 697–698 (2011).
 152. Cermak, T. *et al.* Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Research* **39**, e82–e82 (2011).
 153. Morbitzer, R., Elsaesser, J., Hausner, J. & Lahaye, T. Assembly of custom TALE-type DNA binding domains by modular cloning. *Nucleic Acids Research* **39**, 5790–5799 (2011).
 154. Geißler, R. *et al.* Transcriptional Activators of Human Genes with Programmable DNA-Specificity. *PLoS ONE* **6**, e19509–7 (2011).
 155. Li, L. *et al.* Rapid and highly efficient construction of TALE-based transcriptional regulators and nucleases for genome modification. *Plant Mol Biol* **78**, 407–416 (2012).

156. Sanjana, N. E. *et al.* A transcription activator-like effector toolbox for genome engineering. *Nat Protoc* **7**, 171–192 (2012).
157. Weber, E., Gruetzner, R., Werner, S., Engler, C. & Marillonnet, S. Assembly of Designer TAL Effectors by Golden Gate Cloning. *PLoS ONE* **6**, e19722–5 (2011).
158. Zhang, F., Cong, L., Lodato, S., Kosuri, S. & Church, G. M. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat. Biotechnol.* (2011). doi:10.1038/nbt1775
159. Schmid-Burgk, J. L., Schmidt, T., Kaiser, V., Höning, K. & Hornung, V. A ligation-independent cloning technique for high-throughput assembly of transcription activator-like effector genes. *Nat. Biotechnol.* **31**, 76–81 (2013).
160. Briggs, A. W. *et al.* Iterative capped assembly: rapid and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers. *Nucleic Acids Research* **40**, e117–e117 (2012).
161. Reyon, D. *et al.* FLASH assembly of TALENs for high-throughput genome editing. *Nat. Biotechnol.* **30**, 460–465 (2012).
162. Wang, Z. *et al.* An Integrated Chip for the High-Throughput Synthesis of Transcription Activator-like Effectors. *Angew. Chem. Int. Ed.* **51**, 8505–8508 (2012).
163. Hockemeyer, D. *et al.* Genetic engineering of human pluripotent cells using TALE nucleases. *Nat. Biotechnol.* **29**, 731–734 (2011).
164. Garg, A., Lohmueller, J. J., Silver, P. A. & Armel, T. Z. Engineering synthetic TAL effectors with orthogonal target sites. *Nucleic Acids Research* **40**, 7584–7595 (2012).
165. Li, Y., Moore, R., Guinn, M. & Bleris, L. Transcription activator-like effector hybrids for conditional control and rewiring of chromosomal transgene expression. *Sci. Rep.* **2**, 1–7 (2012).
166. Tremblay, J. P., Chapdelaine, P., Coulombe, Z. & Rousseau, J. Transcription Activator-Like Effector Proteins Induce the Expression of the Frataxin Gene. *Human Gene Therapy* **23**, 883–890 (2012).
167. Mahfouz, M. M. *et al.* Targeted transcriptional repression using a chimeric TALE-SRDX repressor protein. *Plant Mol Biol* **78**, 311–321 (2011).
168. Politz, M. C., Copeland, M. F. & Pflieger, B. F. Artificial repressors for controlling gene expression in bacteria. *Chem. Commun.* **49**, 4325–4327 (2013).
169. Amabile, A. *et al.* Inheritable Silencing of Endogenous Genes by Hit- and-Run Targeted Epigenetic Editing. *Cell* **167**, 219–224.e14 (2016).
170. Mercer, A. C., Gaj, T., Fuller, R. P. & Barbas, C. F. Chimeric TALE recombinases with programmable DNA sequence specificity. *Nucleic Acids Research* **40**, 11163–11172 (2012).
171. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakata, A. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J. Bacteriol.* **169**, 5429–5433 (1987).
172. Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Comparative genomics of defense systems in archaea and bacteria. *Nucleic Acids Research* **41**, 4360–4377 (2013).
173. Pourcel, C., Salvignol, G. & Vergnaud, G. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* **151**, 653–663 (2005).
174. Ehrlich, S. D., Bolotin, A., Quinquis, B. & Sorokin, A. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**, 2551–2561 (2005).
175. Mojica, F. J. M., Díez-Villaseñor, C. S., García-Martínez, J. S. & Soria, E. Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. *J Mol Evol* **60**, 174–182 (2005).
176. Haft, D. H., Selengut, J., Mongodin, E. F. & Nelson, K. E. A Guild of 45 CRISPR-Associated

- (Cas) Protein Families and Multiple CRISPR/Cas Subtypes Exist in Prokaryotic Genomes. *PLoS Comp Biol* **1**, e60–10 (2005).
177. Barrangou, R. *et al.* CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science* **315**, 1709–1712 (2007).
 178. Brouns, S. J. J. *et al.* Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science* **321**, 960–964 (2008).
 179. Marraffini, L. A. & Sontheimer, E. J. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**, 1843–1845 (2008).
 180. Makarova, K. S. *et al.* Evolution and classification of the CRISPR-Cas systems. *Nat Rev Micro* **9**, 467–477 (2011).
 181. van der Oost, J., Westra, E. R., Jackson, R. N. & Wiedenheft, B. Unravelling the structural and mechanistic basis of CRISPR–Cas systems. *Nat Rev Micro* **12**, 479–492 (2014).
 182. Yosef, I., Goren, M. G. & Qimron, U. Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Research* **40**, 5569–5576 (2012).
 183. Swarts, D. C., Mosterd, C., van Passel, M. W. J. & Brouns, S. J. J. CRISPR Interference Directs Strand Specific Spacer Acquisition. *PLoS ONE* **7**, e35888–7 (2012).
 184. Babu, M. *et al.* A dual function of the CRISPR-Cas system in bacterial antiviral immunity and DNA repair. *Molecular Microbiology* **79**, 484–502 (2010).
 185. Plagens, A., Tjaden, B., Hagemann, A., Randau, L. & Hensel, R. Characterization of the CRISPR/Cas Subtype I-A System of the Hyperthermophilic Crenarchaeon *Thermoproteus tenax*. *J. Bacteriol.* **194**, 2491–2500 (2012).
 186. Richter, C., Gristwood, T., Clulow, J. S. & Fineran, P. C. In Vivo Protein Interactions and Complex Formation in the *Pectobacterium atrosepticum* Subtype I-F CRISPR/Cas System. *PLoS ONE* **7**, e49549–11 (2012).
 187. Deveau, H. *et al.* Phage Response to CRISPR-Encoded Resistance in *Streptococcus thermophilus*. *J. Bacteriol.* **190**, 1390–1400 (2008).
 188. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733–740 (2009).
 189. Jinek, M. *et al.* A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816–821 (2012).
 190. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **109**, E2579–86 (2012).
 191. Sinkunas, T. *et al.* In vitro reconstitution of Cascade-mediated CRISPR immunity in *Streptococcus thermophilus*. *The EMBO Journal* 1–10 (2013). doi:10.1038/emboj.2012.352
 192. Semenova, E. *et al.* Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 10098–10103 (2011).
 193. Westra, E. R. *et al.* Type I-E CRISPR-Cas Systems Discriminate Target from Non-Target DNA through Base Pairing-Independent PAM Recognition. *PLoS Genetics* **9**, e1003742–14 (2013).
 194. Westra, E. R. *et al.* CRISPR Immunity Relies on the Consecutive Binding and Degradation of Negatively Supercoiled Invader DNA by Cascade and Cas3. *Molecular Cell* **46**, 595–605 (2012).
 195. Marraffini, L. A. & Sontheimer, E. J. Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* **463**, 568–571 (2010).
 196. Reeks, J., Naismith, J. H. & White, M. F. CRISPR interference: a structural perspective. *Biochem. J.* **453**, 155–166 (2013).
 197. Niewoehner, O., Jinek, M. & Doudna, J. A. Evolution of CRISPR RNA recognition and processing by Cas6 endonucleases. *Nucleic Acids Research* **42**, 1341–1353 (2014).

198. Hatoum-Aslan, A., Samai, P., Maniv, I., Jiang, W. & Marraffini, L. A. A Ruler Protein in a Complex for Antiviral Defense Determines the Length of Small Interfering CRISPR RNAs. *Journal of Biological Chemistry* **288**, 27888–27897 (2013).
199. Zetsche, B. *et al.* Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* 1–14 (2015). doi:10.1016/j.cell.2015.09.038
200. Staals, R. H. J. *et al.* Structure and Activity of the RNA-Targeting Type III-B CRISPR-Cas Complex of *Thermus thermophilus*. *Molecular Cell* **52**, 135–145 (2013).
201. Rouillon, C. *et al.* Structure of the CRISPR Interference Complex CSM Reveals Key Similarities with Cascade. *Molecular Cell* **52**, 124–134 (2013).
202. Spilman, M. *et al.* Structure of an RNA Silencing Complex of the CRISPR-Cas Immune System. *Molecular Cell* **52**, 146–152 (2013).
203. Jore, M. M. *et al.* Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat Struct Mol Biol* **18**, 529–536 (2011).
204. Wiedenheft, B. *et al.* RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 10092–10097 (2011).
205. Nam, K. H. *et al.* Cas5d Protein Processes Pre-crRNA and Assembles into a Cascade-like Interference Complex in Subtype I-C/Dvulg CRISPR-Cas System. *Structure/Folding and Design* **20**, 1574–1584 (2012).
206. Sashital, D. G., Wiedenheft, B. & Doudna, J. A. Mechanism of Foreign DNA Selection in a Bacterial Adaptive Immune System. *Molecular Cell* **46**, 606–615 (2012).
207. Westra, E. R. *et al.* Cascade-mediated binding and bending of negatively supercoiled DNA. *RNA Biology* **9**, 1134–1138 (2014).
208. Wiedenheft, B. *et al.* Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature* **477**, 486–489 (2011).
209. Deng, L., Garrett, R. A., Shah, S. A., Peng, X. & She, Q. A novel interference mechanism by a type IIIB CRISPR-Cmr module in *Sulfolobus*. *Molecular Microbiology* **87**, 1088–1099 (2013).
210. Hale, C. R. *et al.* RNA-Guided RNA Cleavage by a CRISPR RNA-Cas Protein Complex. *Cell* **139**, 945–956 (2009).
211. Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602–607 (2011).
212. Sapranauskas, R. *et al.* The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Research* **39**, 9275–9282 (2011).
213. Garneau, J. E. *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67–71 (2010).
214. Szczelkun, M. D. *et al.* Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 9798–9803 (2014).
215. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* 1–17 (2014). doi:10.1038/nature13011
216. Chylinski, K., Makarova, K. S., Charpentier, E. & Koonin, E. V. Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Research* **42**, 6091–6105 (2014).
217. Jinek, M. *et al.* Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343**, 1247997–1247997 (2014).
218. Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935–949 (2014).
219. Nishimasu, H. *et al.* Crystal Structure of *Staphylococcus aureus* Cas9. *Cell* **162**, 1113–1126 (2015).
220. Jiang, F. & Doudna, J. A. The structural biology of CRISPR-Cas systems. *Curr. Opin. Struct. Biol.* **30**, 100–111 (2015).
221. Anders, C. *et al.* Structural basis of PAM-dependent target DNA recognition by the Cas9

- endonuclease. *Nature* **513**, 569–573 (2014).
222. Sternberg, S. H., LaFrance, B., Kaplan, M. & Doudna, J. A. Conformational control of DNA target cleavage by CRISPR–Cas9. *Nature* **527**, 110–113 (2015).
223. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
224. Mali, P. *et al.* RNA-Guided Human Genome Engineering via Cas9. *Science* **339**, 823–826 (2013).
225. Jinek, M. *et al.* RNA-programmed genome editing in human cells. *eLife* **2**, 273–9 (2013).
226. Mussolino, C. & Cathomen, T. RNA guides genome engineering. *Nat. Biotechnol.* (2013). doi:10.1038/nbt.2527
227. Wang, H. *et al.* One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. *Cell* 1–9 (2013). doi:10.1016/j.cell.2013.04.025
228. Li, W., Teng, F., Li, T. & Zhou, Q. Simultaneous generation and germline transmission of multiple gene mutations in rat using CRISPR-Cas systems. *Nature Publishing Group* **31**, 684–686 (2013).
229. Jao, L. E., Wente, S. R. & Chen, W. Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. in (2013). doi:10.1073/pnas.1308335110/-/DCSupplemental/pnas.201308335SI.pdf
230. Hsu, P. D., Lander, E. S. & Zhang, F. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell* **157**, 1262–1278 (2014).
231. Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* **32**, 347–355 (2014).
232. Mali, P., Esvelt, K. M. & Church, G. M. Cas9 as a versatile tool for engineering biology. *Nat. Methods* **10**, 957–963 (2013).
233. Choi, P. S. & Meyerson, M. Targeted genomic rearrangements using CRISPR/Cas technology. *Nat Comms* **5**, 1–6 (1AD).
234. Torres, R. *et al.* Engineering human tumour-associated chromosomal translocations with the RNA-guided CRISPR-Cas9 system. *Nat Comms* **5**, 3964 (2014).
235. Xue, W. *et al.* CRISPR-mediated direct mutation of cancer genes in the mouse liver. *Nature* 1–16 (2014). doi:10.1038/nature13589
236. Heckl, D. *et al.* Generation of mouse models of myeloid malignancy with combinatorial genetic lesions using CRISPR-Cas9 genome editing. *Nat. Biotechnol.* **32**, 941–946 (2014).
237. Schwank, G. *et al.* Functional Repair of CFTR by CRISPR/Cas9 in Intestinal Stem Cell Organoids of Cystic Fibrosis Patients. *Stem Cell* **13**, 653–658 (2013).
238. Wu, Y. *et al.* Correction of a Genetic Disease in Mouse via Use of CRISPR-Cas9. *Stem Cell* **13**, 659–662 (2013).
239. Wang, F. & Qi, L. S. Applications of CRISPR Genome Engineering in Cell Biology. *Trends Cell Biol.* 1–14 (2016). doi:10.1016/j.tcb.2016.08.004
240. Yang, H. *et al.* One-Step Generation of Mice Carrying Reporter and Conditional Alleles by CRISPR/Cas-Mediated Genome Engineering. *Cell* 1–10 (2013). doi:10.1016/j.cell.2013.08.022
241. Niu, Y. *et al.* Generation of Gene-Modified Cynomolgus Monkey via Cas9/RNA-Mediated Gene Targeting in One-Cell Embryos. *Cell* 1–8 (2014). doi:10.1016/j.cell.2014.01.027
242. Doudna, J. A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096–1258096 (2014).
243. Zhang, H. *et al.* The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation. *Plant Biotechnol J* **12**, 797–807 (2014).
244. Waltz, E. Gene-edited CRISPR mushroom escapes US regulation. *Nature* **532**, 293–293 (2016).
245. Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–87 (2014).

246. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80–84 (2014).
247. Zhou, Y. *et al.* High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature* 1–16 (2014). doi:10.1038/nature13166
248. Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315–326 (2006).
249. Chen, S. *et al.* Genome-wide CRISPR Screen in a Mouse Model of Tumor Growth and Metastasis. *Cell* 1–16 (2015). doi:10.1016/j.cell.2015.02.038
250. Parnas, O. *et al.* A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks. *Cell* 1–13 (2015). doi:10.1016/j.cell.2015.06.059
251. Wong, A. S. L. *et al.* Multiplexed barcoded CRISPR-Cas9 screening enabled by CombiGEM. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 2544–2549 (2016).
252. Evers, B. *et al.* CRISPR knockout screening outperforms shRNA and CRISPRi in identifying essential genes. *Nat. Biotechnol.* **34**, 631–633 (2016).
253. Morgens, D. W., Deans, R. M., Li, A. & Bassik, M. C. Systematic comparison of CRISPR/Cas9 and RNAi screens for essential genes. *Nature Publishing Group* 1–4 (2016). doi:10.1038/nbt.3567
254. Gilbert, L. A. *et al.* CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. *Cell* 1–10 (2013). doi:10.1016/j.cell.2013.06.044
255. Gilbert, L. A. *et al.* Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* 1–15 (2014). doi:10.1016/j.cell.2014.09.029
256. Larson, M. H. *et al.* CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat Protoc* **8**, 2180–2196 (2013).
257. Zhao, Y. *et al.* Sequence-specific inhibition of microRNA via CRISPR/CRISPRi system. *Sci. Rep.* **4**, 1–5 (2014).
258. Perez-Pinera, P. *et al.* RNA-guided gene activation by CRISPR-Cas9–based transcription factors. *Nat. Methods* **10**, 973–976 (2013).
259. Cheng, A. W. *et al.* Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Research* 1–9 (2013). doi:10.1038/cr.2013.122
260. Chavez, A. *et al.* Comparison of Cas9 activators in multiple species. *Nat. Methods* **13**, 563–567 (2016).
261. 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 (2014).
262. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* 1–18 (2014). doi:10.1038/nature14136
263. Chavez, A. *et al.* Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods* **12**, 326–328 (2015).
264. Zalatan, J. G. *et al.* Engineering Complex Synthetic Transcriptional Programs with CRISPR RNA Scaffolds. *Cell* 1–12 (2014). doi:10.1016/j.cell.2014.11.052
265. Hilton, I. B. *et al.* Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nature Publishing Group* 1–10 (2015). doi:10.1038/nbt.3199
266. Kearns, N. A. *et al.* Functional annotation of native enhancers with a Cas9-histone demethylase fusion. *Nat. Methods* **12**, 401–403 (2015).
267. Vojta, A. *et al.* Repurposing the CRISPR-Cas9 system for targeted DNA methylation. *Nucleic Acids Research* **44**, 5615–5628 (2016).
268. Chen, B. *et al.* Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System. *Cell* **155**, 1479–1491 (2013).
269. Anton, T., Bultmann, S., Leonhardt, H. & Markaki, Y. Visualization of specific DNA sequences in living mouse embryonic stem cells with a programmable fluorescent

- CRISPR/Cas system. *Nucleus* **5**, 163–172 (2014).
270. Ma, H. *et al.* Multicolor CRISPR labeling of chromosomal loci in human cells. *Proceedings of the National Academy of Sciences* **112**, 3002–3007 (2015).
271. Deng, W., Shi, X., Tjian, R., Lionnet, T. & Singer, R. H. CASFISH: CRISPR/Cas9-mediated in situ labeling of genomic loci in fixed cells. *Proceedings of the National Academy of Sciences* **112**, 11870–11875 (2015).
272. Nelles, D. A. *et al.* Programmable RNA Tracking in Live Cells with CRISPR/Cas9. *Cell* 1–10 (2016). doi:10.1016/j.cell.2016.02.054
273. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
274. Nishida, K. *et al.* Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**, (2016).
275. Senís, E. *et al.* CRISPR/Cas9-mediated genome engineering: An adeno-associated viral (AAV) vector toolbox. *Biotechnology Journal* **9**, 1402–1412 (2014).
276. Müller, M. *et al.* Streptococcus thermophilus CRISPR-Cas9 Systems Enable Specific Editing of the Human Genome. *Molecular Therapy* **24**, 636–644 (2015).
277. 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 (2013).
278. Ran, F. A. *et al.* In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature* **520**, 186–191 (2015).
279. Yang, L. *et al.* Optimization of scarless human stem cell genome editing. *Nucleic Acids Research* **41**, 9049–9061 (2013).
280. Kleinstiver, B. P. *et al.* Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* 1–17 (2015). doi:10.1038/nature14592
281. 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 (2016).
282. Kleinstiver, B. P. *et al.* Broadening the targeting range of *Staphylococcus aureus* CRISPR-Cas9 by modifying PAM recognition. *Nat. Biotechnol.* **33**, 1293–1298 (2015).
283. Leenay, R. T. *et al.* Identifying and Visualizing Functional PAM Diversity across CRISPR-Cas Systems. *Molecular Cell* 1–12 (2016). doi:10.1016/j.molcel.2016.02.031
284. Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR–Cas systems. *Nat Rev Micro* **13**, 722–736 (2015).
285. Yamano, T. *et al.* Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA. *Cell* **165**, 949–962 (2016).
286. Dong, D. *et al.* The crystal structure of Cpf1 in complex with CRISPR RNA. *Nature* **532**, 522–526 (2016).
287. Maresca, M., Lin, V. G., Guo, N. & Yang, Y. Obligate Ligation-Gated Recombination (ObLiGaRe): Custom-designed nuclease-mediated targeted integration through nonhomologous end joining. *Genome Research* **23**, 539–546 (2013).
288. Hur, J. K. *et al.* Targeted mutagenesis in mice by electroporation of Cpf1 ribonucleoproteins. *Nat. Biotechnol.* **34**, 807–808 (2016).
289. Kim, Y. *et al.* Generation of knockout mice by Cpf1-mediated gene targeting. *Nat. Biotechnol.* **34**, 808–810 (2016).
290. Abudayyeh, O. O. *et al.* C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **353**, aaf5573 (2016).
291. Gao, F., Shen, X. Z., Jiang, F., Wu, Y. & Han, C. DNA-guided genome editing using the *Natronobacterium gregoryi* Argonaute. *Nature Publishing Group* **34**, 768–773 (2016).
292. Cyranoski, D. Replications, ridicule and a recluse: the controversy over NgAgo gene-editing intensifies. *Nature* **536**, 136–137 (2016).
293. Koo, T., Lee, J. & Kim, J.-S. Measuring and Reducing Off-Target Activities of

- Programmable Nucleases Including CRISPR-Cas9. *Molecules and Cells* **38**, 475–481 (2015).
294. Lee, H. J., Kweon, J., Kim, E., Kim, S. & Kim, J. S. Targeted chromosomal duplications and inversions in the human genome using zinc finger nucleases. *Genome Research* **22**, 539–548 (2012).
295. Pattanayak, V., Ramirez, C. L., Joungh, J. K. & Liu, D. R. Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nat. Methods* **8**, 765–770 (2011).
296. Guilinger, J. P. *et al.* Broad specificity profiling of TALENs results in engineered nucleases with improved DNA-cleavage specificity. *Nat. Methods* **11**, 429–435 (2014).
297. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31**, 233–239 (2013).
298. Cho, S. W. *et al.* Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Research* **24**, 132–141 (2014).
299. Cradick, T. J., Fine, E. J., Antico, C. J. & Bao, G. CRISPR/Cas9 systems targeting β -globin and CCR5 genes have substantial off-target activity. *Nucleic Acids Research* **41**, 9584–9592 (2013).
300. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* 1–8 (2013). doi:10.1038/nbt.2647
301. Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* 1–6 (2013). doi:10.1038/nbt.2623
302. Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.* **31**, 839–843 (2013).
303. Lin, Y. *et al.* CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucleic Acids Research* **42**, 7473–7485 (2014).
304. 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 (2015).
305. Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**, 187–197 (2014).
306. Yee, J.-K. Off-target effects of engineered nucleases. *FEBS J* **283**, 3239–3248 (2016).
307. Veres, A. *et al.* Low Incidence of Off-Target Mutations in Individual CRISPR-Cas9 and TALEN Targeted Human Stem Cell Clones Detected by Whole-Genome Sequencing. *Stem Cell* **15**, 27–30 (2014).
308. Smith, C. *et al.* Whole-Genome Sequencing Analysis Reveals High Specificity of CRISPR/Cas9 and TALEN-Based Genome Editing in Human iPSCs. *Stem Cell* **15**, 12–13 (2014).
309. Vouillot, L., Th  lie, A. & Pollet, N. Comparison of T7E1 and surveyor mismatch cleavage assays to detect mutations triggered by engineered nucleases. *G3 (Bethesda)* **5**, 407–415 (2015).
310. Brinkman, E. K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Research* **42**, e168–e168 (2014).
311. Wu, X. *et al.* Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat. Biotechnol.* **32**, 670–676 (2014).
312. Kuscu, C., Arslan, S., Singh, R., Thorpe, J. & Adli, M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat. Biotechnol.* **32**, 1–9 (2014).
313. Gabriel, R. *et al.* An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat. Biotechnol.* **29**, 816–823 (2011).
314. Hu, J. *et al.* Detecting DNA double-stranded breaks in mammalian genomes by linear amplification-mediated high-throughput genome-wide translocation sequencing. *Nat*

- Protoc* **11**, 853–871 (2016).
315. Frock, R. L. *et al.* Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nature Publishing Group* 1–10 (2014). doi:10.1038/nbt.3101
 316. Crosetto, N. *et al.* Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. *Nat. Methods* **10**, 361–365 (2013).
 317. Kim, D. *et al.* Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat. Methods* 1–8 (2015). doi:10.1038/nmeth.3284
 318. Isaac, R. S. *et al.* Nucleosome breathing and remodeling constrain CRISPR-Cas9 function. *eLife* **5**, 1 (2016).
 319. Horlbeck, M. A. *et al.* Nucleosomes impede Cas9 access to DNA in vivo and in vitro. *eLife* **5**, 2767 (2016).
 320. Kleinstiver, B. P. *et al.* Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. *Nat. Biotechnol.* **34**, 869–874 (2016).
 321. Kim, D. *et al.* Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nat. Biotechnol.* **34**, 863–868 (2016).
 322. 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 (2014).
 323. Mali, P. *et al.* CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* **31**, 833–838 (2013).
 324. Ran, F. A. *et al.* Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell* **154**, 1380–1389 (2013).
 325. Shen, B. *et al.* Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat. Methods* **11**, 399–402 (2014).
 326. Kim, E. *et al.* Precision genome engineering with programmable DNA-nicking enzymes. *Genome Research* **22**, 1327–1333 (2012).
 327. Tsai, S. Q. *et al.* Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat. Biotechnol.* **32**, 569–576 (2014).
 328. Gabsalilow, L., Schierling, B., Friedhoff, P., Pingoud, A. & Wende, W. Site- and strand-specific nicking of DNA by fusion proteins derived from MutH and I-SceI or TALE repeats. *Nucleic Acids Research* **41**, e83–e83 (2013).
 329. 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 (2014).
 330. 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. *Human Gene Therapy* **26**, 425–431 (2015).
 331. Bolukbasi, M. F. *et al.* DNA-binding-domain fusions enhance the targeting range and precision of Cas9. *Nat. Methods* **12**, 1150–1156 (2015).
 332. 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 Research* **24**, 1012–1019 (2014).
 333. Zuris, J. A. *et al.* Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nature Publishing Group* **33**, 73–80 (2015).
 334. Ramakrishna, S. *et al.* Gene disruption by cell-penetrating peptide-mediated delivery of Cas9 protein and guide RNA. *Genome Research* **24**, 1020–1027 (2014).
 335. Gaj, T., Guo, J., Kato, Y., Sirk, S. J. & Barbas, C. F. Targeted gene knockout by direct delivery of zinc-finger nuclease proteins. *Nat. Methods* **9**, 805–807 (2012).
 336. Flisikowska, T. *et al.* Efficient Immunoglobulin Gene Disruption and Targeted Replacement in Rabbit Using Zinc Finger Nucleases. *PLoS ONE* **6**, e21045 (2011).
 337. Liu, J., Gaj, T., Patterson, J. T., Sirk, S. J. & Barbas, C. F., III. Cell-Penetrating Peptide-Mediated Delivery of TALEN Proteins via Bioconjugation for Genome Engineering. *PLoS ONE* **9**, e85755–7 (2014).

338. Tesson, L. *et al.* Knockout rats generated by embryo microinjection of TALENs. *Nat. Biotechnol.* **29**, 695–696 (2011).
339. Dow, L. E. *et al.* Inducible in vivo genome editing with CRISPR-Cas9. *Nat. Biotechnol.* **33**, 390–394 (2015).
340. González, F. *et al.* An iCRISPR platform for rapid, multiplexable, and inducible genome editing in human pluripotent stem cells. *Cell Stem Cell* **15**, 215–226 (2014).
341. Nuñez, J. K., Harrington, L. B. & Doudna, J. A. Chemical and Biophysical Modulation of Cas9 for Tunable Genome Engineering. *ACS Chem. Biol.* **11**, 681–688 (2016).
342. Davis, K. M., Pattanayak, V., Thompson, D. B., Zuris, J. A. & Liu, D. R. Small molecule-triggered Cas9 protein with improved genome-editing specificity. *Nat. Chem. Biol.* **11**, 316–318 (2015).
343. Zetsche, B., Volz, S. E. & Zhang, F. A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nat. Biotechnol.* **33**, 139–142 (2015).
344. Nihongaki, Y., Kawano, F., Nakajima, T. & Sato, M. Photoactivatable CRISPR-Cas9 for optogenetic genome editing. *Nat. Biotechnol.* **33**, 755–760 (2015).
345. Truong, D.-J. J. *et al.* Development of an intein-mediated split-Cas9 system for gene therapy. *Nucleic Acids Research* **43**, 6450–6458 (2015).
346. Hemphill, J., Borchardt, E. K., Brown, K., Asokan, A. & Deiters, A. Optical Control of CRISPR/Cas9 Gene Editing. *J. Am. Chem. Soc.* **137**, 5642–5645 (2015).
347. Kleinstiver, B. P. *et al.* High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490–495 (2016).
348. Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science* 1–7 (2015). doi:10.1126/science.aad5227
349. DiCarlo, J. E. *et al.* Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Research* **41**, 4336–4343 (2013).
350. Jegga, A. G., Inga, A., Menendez, D., Aronow, B. J. & Resnick, M. A. Functional evolution of the p53 regulatory network through its target response elements. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 944–949 (2008).
351. Tomso, D. J. *et al.* Functionally distinct polymorphic sequences in the human genome that are targets for p53 transactivation. *Proceedings of the National Academy of Sciences* **102**, 6431–6436 (2005).
352. Stuckey, S. & Storici, F. Gene knockouts, in vivo site-directed mutagenesis and other modifications using the delitto perfetto system in *Saccharomyces cerevisiae*. *Meth. Enzymol.* **533**, 103–131 (2013).
353. Geissmann, Q. OpenCFU, a New Free and Open-Source Software to Count Cell Colonies and Other Circular Objects. *PLoS ONE* **8**, e54072–10 (2013).
354. Pizzato, M. *et al.* A one-step SYBR Green I-based product-enhanced reverse transcriptase assay for the quantitation of retroviruses in cell culture supernatants. *Journal of Virological Methods* **156**, 1–7 (2009).
355. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
356. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
357. 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**, 483–483 (2016).
358. Petris, G., Bestagno, M., Arnoldi, F. & Burrone, O. R. New tags for recombinant protein detection and O-glycosylation reporters. *PLoS ONE* **9**, e96700 (2014).
359. Campeau, E. *et al.* A Versatile Viral System for Expression and Depletion of Proteins in Mammalian Cells. *PLoS ONE* **4**, e6529–18 (2009).
360. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. - PubMed - NCBI. *Nat. Methods* **11**, 783–784 (2014).
361. Dang, Y. *et al.* Optimizing sgRNA structure to improve CRISPR-Cas9 knockout efficiency.

- Genome Biol* 1–10 (2015). doi:10.1186/s13059-015-0846-3
362. Nelson, C. E. & Gersbach, C. A. Cas9 loosens its grip on off-target sites. *Nat. Biotechnol.* **34**, 298–299 (2016).
363. Tsai, S. Q. & Joung, J. K. Defining and improving the genome-wide specificities of CRISPR-Cas9 nucleases. *Nat. Rev. Genet.* **17**, 300–312 (2016).
364. Paquet, D. *et al.* Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature* **533**, 125–129 (2016).
365. Chick, H. E. *et al.* Integrase-deficient lentiviral vectors mediate efficient gene transfer to human vascular smooth muscle cells with minimal genotoxic risk. *Human Gene Therapy* **23**, 1247–1257 (2012).
366. Moore, R. *et al.* CRISPR-based self-cleaving mechanism for controllable gene delivery in human cells. *Nucleic Acids Research* **43**, 1297–1303 (2015).
367. Kiani, S. *et al.* Cas9 gRNA engineering for genome editing, activation and repression. *Nature Publishing Group* 1–6 (2015). doi:10.1038/nmeth.3580
368. Wright, A. V. *et al.* Rational design of a split-Cas9 enzyme complex. *Proceedings of the National Academy of Sciences* **112**, 2984–2989 (2015).
369. Chew, W. L. *et al.* A multifunctional AAV-CRISPR-Cas9 and its host response. *Nat. Methods* **13**, 868–874 (2016).
370. Cox, D. B. T., Platt, R. J. & Zhang, F. Therapeutic genome editing: prospects and challenges. *Nat Med* **21**, 121–131 (2015).

Appendix

- **High-fidelity SpCas9 variants related oligos**

Table 1. Primers used for plasmid cloning

Primer name	Sequence (5' ... -3')
Rec1-II-NheI-F	CCAGAAAGCACAAGTTGCTAGCCAGGGGGACAGTC
Rec1-II-NheI-R	GACTGTCCCCCTGGCTAGCAACTTGTGCTTTCTGG
Rec1-II-NcoI-F	CAGCGCACTTTCGACCATGGAAGCATCCCCCA
Rec1-II-NcoI-R	TGGGGGATGCTTCCATGTCGAAAGTGCGCTG
T3-Forward	AATTAACCCTCACTAAAGGG
T7-Reverse	TAATACGACTCACTATAGGG
sgRNA-Ontarget-F	CTCGTGACCACCCTGACCTAGTTTTAGAGCTAGAAATAGCAA
sgRNA-Ontarget-R	TAGGTCAGGGTGGTCACGAGGATCATTTATCTTTCACTGCG
Apa-ZhangCas-F	ACGTGGGCCCTCTGGCCAG
Nhe-ZhangCas-R	TACGCTAGCTCCCTTTTTCTTTTTGCCTGG
Apa-JoungCas-F	ATTAGGGCCCCCTGGCCCGAGGGAAC
Spe-JoungCas-R	TAATACTAGTGACTTTCCTCTTCTTCTGGG

Table 2. Primers used yeast reporter cassette construction

Primer name	Sequence (5' ... -3')
TRP1-genomic-F	CCAAGAGGGAGGGCATTGG
TRP1Pt1-ON-Kpn-R	TGCGGTACCGTAGGTCAGGGTGGTCACGAGTTAGAGGAACTCTTGG TATTCTTGC
TRP1Pt2-Kpn-F	TTAGGTACCGTAATCAACCTAAGGAGGATGTTT
TRP1-genomic-R	TGCTTGCTTTTCAAAGGCCTG
ADE2-genomic-F	TGCCTAGTTTCATGAAATTTAAAGC
ADE2Pt1-OFF1-Bam-R	CCAGGATCCGGAGGTCAGGGTGGTCACGAGTTAGACGCAAGCATCA ATGGTAT
ADE2Pt1-OFF2-Bam-R	CCAGGATCCGTAGGTAAGGGTGGTCACGAGTTAGACGCAAGCATCA ATGGTAT
ADE2Pt1-OFF3-Bam-R	CCAGGATCCGTAGGTCAGGGCGGTCACGAGTTAGACGCAAGCATCA ATGGTAT
ADE2Pt1-OFF4-Bam-R	CCAGGATCCGTAGGTCAGGGTGGTAACGAGTTAGACGCAAGCATCA ATGGTAT
ADE2Pt2-Bam-F	ATTAGGATCCTGGTGTGGAAATGTTCTATTTAG
ADE2-genomic-R	GTAATCATAACAAAGCCTAAAAATAG
TRP1-CORE-F	TATTGAGCACGTGAGTATACGTGATTAAGCACACAAAGGCAGCTTGG AGTTAGGGATAACAGGGTAATTTGGATGGACGCAAAGAAGT
TRP1-CORE-R	TGCAGGCAAGTGCACAAACAATACTTAATAAATACTACTCAGTAATA ACTTCGTACGCTGCAGGTCGAC
ADE2-CORE-F	CCTACTATAACAATCAAGAAAAACAAGAAAATCGGACAAAACAATCA AGTTAGGGATAACAGGGTAATTTGGATGGACGCAAAGAAGT

ADE2-CORE-R	ATATCATTTTATAATTATTTGCTGTACAAGTATATCAATAAACTTATAT ATTCGTACGCTGCAGGTCGAC
-------------	--

Table 3. On and off target sites used to generate the yACMO strains

Target name	Sequence (5'- ... -3', with lowercase PAM), mismatch in bold
TRP1-on	CTCGTGACCACCCTGACCTAcgg
ADE2-off1	CTCGTGACCACCCTGACCT C cg
ADE2-off2	CTCGTGACCACCCT T ACCTAcgg
ADE2-off3	CTCGTGACC G CCCTGACCTAcgg
ADE2-off4	CTCGT T ACCACCCTGACCTAcgg

Table 4. Error-prone PCR primers

Primer name	Sequence (5'- ... -3')
epPCR-F	GTCTAAAAATGGCTACGCCGGATACATTGACGGCGGAGCAAGCCAG GAGG
epPCR-R	TCTCGGGCCATCTCGATAACGATATTCTCGGGCTTATGCCTTCCCATT AC

Table 5. Spacer sequences used to prepare sgRNAs for reporter assays

Target name	Spacer sequence (5'- ... -3')
GFPon	GGGCACGGGCAGCTTGCCGG
GFP1314	GGGCAC CC GCAGCTTGCCGG
GFP1819	GCC CACGGGCAGCTTGCCGG
GFP18	GGCC CACGGGCAGCTTGCCGG
GFP site 2	GTCGCCCTCGAACTTCACT
GFP site 14	GAAGGGCATCGACTTCAAGG
GFP site 16	GCTGAAGCACTGCACGCCGT
GFP site 18	GACCAGGATGGGCACCACCC
GFP site 20	GAAGTTCGAGGGCGACACCC
GFPon-19nt	GGCACGGGCAGCTTGCCGG
GFPB-18nt	GGCAAGCTGCCCCGTGCC
GFPW-17nt	GTGACCACCCTGACCTA
GFP site 20 (+G)	gGAAGTTCGAGGGCGACACCC
GFP site 25 (+G)	gCCTCGAACTTCACTCGGCG
GFP site 25 (19+G)	gCTCGAACTTCACTCGGCG
GFP site 25 (+2+G)	gGCCCTCGAACTTCACTCGGCG
GFP site 25 (+3+G)	gCGCCCTCGAACTTCACTCGGCG
GFP site 25 (+2)	GCCCTCGAACTTCACTCGGCG
GFP 5'-C	CTCGTGACCACCCTGACCTA
GFP 5'-C+G	gCTCGTGACCACCCTGACCTA
GFP 5'-T	TTCAAGTCCGCCATGCCCGA
GFP 5'-T+G	gTTCAAGTCCGCCATGCCCGA
GFP 5'-A	ACCATCTTCTTCAAGGACGA
GFP 5'-A+G	gACCATCTTCTTCAAGGACGA

TetO-on	GTGATAGAGAACGTATGTCTG
TetO-on+G	gGTGATAGAGAACGTATGTCTG
TetO-off6	GTGATAGAGAACGT CT GTCTG
TetO-off1314	GTGATA CT GAAACGTATGTCTG
TetO-off1819	GAC ATAGAGAACGTATGTCTG

5'- mismatched nucleotides are indicated in lowercase. Mutations are indicated in bold.

Table 6. Targeted deep-sequencing oligos

Common forward overhang: 5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3'

Common reverse overhang: 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3'

Locus	Target	Forward (5'...-3')	Reverse (5'...-3')
EMX1-on	GAGTCCGAGCAGAAGAAGAAgagg	CCGGAGGACAAAGTACAAACGGC	AAGCAGCACTCTGCCCTCGTG
EMX1-ot1	GAGTTAGAGCAGAAGAAGAAgagg	CTTTTATACCATCTTGGGGTTACAG	CTAGGAAAGATTACAGAGAGTCTGAC
EMX1-ot2	GAGTCT A AGCAGAAGAAGAAgagg	CAATGTGCTTCAACCCATCACGGC	CCTCTACTTCATTGTAACAAGGTAAG
EMX1-ot3	A AGTCTGAGCA CA AGAAGAAtgg	TAGTTCTGACATTCCTCCTGAGGG	CTCTGTTGTTATTTTTGGTCAATATCTG
EMX1-ot4	GAGTCCTAGCAG G AGAAGAAGag	AAAGCCTGGAGGCTGCCAGGT	ATCTAGCTGTCTGTCTCATTGGC
EMX1-ot5	GAG G CCGAGCAGAAGA AAG Acgg	CAGGAGCCGGGTGGGAG	CCTCAGCCTTCCCTCAGCCAC
VEGFA3-on	GGTGAGTGAGTGTGTGCGTgagg	CTGGGTGAATGGAGCGAGCAG	GGAAGGCGGAGAGCCGGACA
VEGFA3-ot1	A GTGAGTGAGTGTGTGTgagg	GAAGGGGAGGGGGAAGTCACC	CGTGCGTGCCGCCGTTGATC
VEGFA3-ot2	TGT G GTGAGTGTGTGCGTgagg	TCTGTCAACACACAGTTACCACC	GTAGTTGCTGGGGATGGGGTATG
VEGFA3-ot3	GCTGAGTGAGTGT A TGCGTgagg	CACCTGGCCCATTTCTCCTTTGAGG	TGGGGACAGCATGTGCAAGCCACA
VEGFA3-ot4	GGTGAGTGAGTGTGTGTgagg	GGACCCCTCTGACAGACTGCA	CACACACCCTCACATACCCTCAC
VEGFA3-ot5	A GAGAGTGAGTGTGT C ATgagg	GGAAGAATGCAAAGGAGAAGCAAGTAC	GACCTGGTGGGAGTTGATTGGATC
VEGFA3-ot6	A GTGTGTGAGTGTGTGCGTgagg	CCTTGGGAATCTATCTGAATAGGCCT	GACACCCACACACTCTCATGC
VEGFA3-ot7	TGTGAGT A AGTGTGTGTgagg	CCTAAGCTGTATGTGAGTCCCTGA	CTGTTTTGCTAAGAGATGATTAGATGGTC
VEGFA3-ot8	GTTGAGT GA ATGTGTGCGTgagg	GCCCTCTCCGGAAGTGCCTTG	GAAGGGTTGGTTTGAAGGCTGTC
VEGFA3-ot9	GGTGAGTGAGT G CGTGCGgagg	CCACAGGAATTTGAAGTCCGTGCT	CCCCACGTCCACCCATACACAC
VEGFA3-ot10	A GCAGTG G GTGTGTGCGTgagg	GACGTCTGGGTCCCGAGCAGT	GGCCGTGAGTCCGTCGCCGA
VEGFA3-ot11	TGTGAGTGAGTGTGTGCGTgagg	GGAGGGTTGAACTGTGACAGAACTG	TGAGTATGTGTGAGTGTGAGTGTGCA
VEGFA3-ot12	A CTGTGTGAGTGTGTGCGTgagg	GATCCTTAGGCGTGCCTGTGC	CACCGGCACAGTGTGACTCAC
VEGFA3-ot13	TGTGAGTGAGTGTGT A Tgagg	AGACCTTCAATGTGGATGTGCGTG	CATAGAGTGTAGCAGATTTCCATAACTTC

Mutations are indicated in bold. PAM are included in lowercase.

• SLICES circuits

Table 7. Primers used for lentiSLICES plasmid cloning

Primer name	Sequence (5' ... -3')
Cas9-intron-F	CTTTAAAGAGGACATCCAGAAAGCCAGGTAAGGGGCTCACAGTAG C
Cas9-intron-R	CGTGCAGGCTATCGCCCTGGCCGGACACCTGTGGAGAGAAAGGCAA AG
Xho-CMV-F	ATTACTCGAGGTTGACATTGATTATTGACTAGTTA
CMV-TO-Bam-R	AATGGATCCTCTCCAGGCGATCTGACG

Mlu-EMCV-IRES-F	ATTAACGCGTGTTATTTCCACCATATTGCCG
Blast-Mlu-R	ATTAACGCGTTTAGCCCTCCCACACATAAC
Eco-H1TOgRNA-F	TAATGAATTCTAGTAGAATTGAGGTACCAATATTTGCATGTCGCTATG TG
gRNA-Mfel-R	ATTCAATTGAAAAAGCACCCGACTCGGTGC
Nhe-STOP-Mlu-1	CTAGCTGATAATGTACA
Nhe-STOP-Mlu-2	CGCGTGTACATTATCAG

Table 8. Spacer sequences used to prepare sgRNAs

Target name	Spacer sequence (5' ... -3')
Cas-a	gTACGCCGGCTACATTGACGG
Cas-b	GATCCTTGTAGTCTCCGTCG
Cas-c	GGCTACGCCGGCTACATTGA
GFPW	gCTCGTGACCACCCTGACCTA
GFPM	gCTCGTGACCACCCTGACCT C
GFPMM	gCTCGT C ACCACCCTGACCT C
STh1-1	GGCAGAAGGCTGACCCGGCG
STh1-2	gGCCTACAGAAGCGAGGCC
STh1-3	gAGACTAACGAGGACGACGA
RepSV5	GTCCCCTCCACCCACAGTG

Mismatched 5'-G are indicated in lowercase. Mutations are indicated in bold.

Table 9. Target sequences for *S. thermophilus* NHEJ reporters

Target name	Target sequence (5' ... -3', PAM in lowercase)
NHEJ-Rep.W	GTCCCCTCCACCCACAGTGcaagaaa
NHEJ-Rep.M	GTCCCCTCCACCC A ACAGTGcaagaaa

Mutations are indicated in bold.

- **sgRNAs targeting genomic sites and oligos to amplify each locus**

Table 10. Spacer sequences for genomic targets & oligos for the amplification of genomic loci

Locus	Target	Forward (5'...-3')	Reverse (5'...-3')
GFP	gCTCGTGACCACCCTGACCTACGG	ACCATGGTGAGCAAGGGCGAGGA	AGCTCGTCCATGCCGAGAGTGATC
VEGFA3	GGTGAGTGAGTGTGTGCGTGTGG	GCATACGTGGGCTCCAACAGGT	CCGCAATGAAGGGGAAGCTCGA
VEGFA3-ot1	AG TGAGTGAGTGTGTG TGGG	CAGGCGCCTTGGGCTCCGTCA	CCCAGGATCCGCGGGTCAC
VEGFA3-ot2	TGTGGG TGAGTGTGTGCG TAGG	AGTCAGCCCTCTGTATCCCTGGA	GAGATATCTGCACCCTCATGTTAC
ZSCAN2	GTGCGCAAGAGCTTCAGCCGGG	GACTGTGGCAGAGGTTGAGC	TGTATACGGGACTTGACTCAGACC
ZSCAN2-ot1	GTGTGGCAAG GG CTTCAGCC AGG	CACGACTGCAGGCTCATGAGC	GAAGCGCTTACCACACACATCAC
ZSCAN2-ot2	ATGGGGA AAGAGCTTCAG CCTGG	AGTCACATGCTGCCTGGATTGAC	GTGGAGGAGATTTCTTAGGAGAG
EMX1-K	GAGTCCGAGCAGAAGAAGAAGGG	CTGCCATCCCCTTCTGTGAATGT	GGAATCTACCACCCAGGCTCT
EMX1-R	gGCCTCCCAAAGCCTGGCCAGGG	same as above	same as above
EMX1-K-ot1	GAG T AGAGCAGAAGA AAAGG	TGTGGGAGATTTGCATCTGTGGA	TTGAGACATGGGATAGAATCATGAAC
EMX1-K-ot2	GAG T A AGCAGAAGA AGAG	CTGCTGTTTCTGAAGCTGCCACT	CTGCCATGGAAATTCAGAGGGAAC

EMX1-R-ot1	AC CTCCCCATAGCCTGGCCAGGG	TGAACGAATCAGGTCTGAGAGGATC	GAGCTTCACTCCAGAGAGGCTGT
EMX1-R-ot2	TC CTCCCCACAGCCTGGCCAGGG	TGCTACTGCTGGCTGCAGAGATG	GCATTCGTTTTGGGAGGCAGAGGA
VEGFA2	GACCCCTCCACCCCGCTCCGG	TCAGCGGACTCACCGGCCAG	GCGCCGAGTCGCCACTGCGG
FANCF2	GCTGCAGAAGGGATTCCATGAGG	GCCAGGCTCTTTGGAGTGTC	AGCATAGCGCCTGGCATTAAATAGG
FANCF2-ot1	GCTGCAGAAGGGATTCCA A GAGG	CCCGTGAGGTGCTGAGATTTGAAC	CACATGGAGGAGGTGACGCTG
CCR5sp11	GGTACCTATCGATTGTCAGGAGG	ATGCACAGGGTGAACAAGATGGA	CTAAGCCATGTGCACAACCTGAC
OT-CCR2	GGTATCTATCGATTGTCAGGAGG	CATTGTGGGCTCACTCTGCTGCA	CTGAGATGAGCTTTCTGGAGAGTCA
CXCR4	GGAAGCGTGATGACAAAGAGG	AGAGGAGTTAGCCAAGATGTGACTTTGAAACC	GGACAGGATGACAATACCAGGCAGGATAAGGCC

Mismatched 5'-G are indicated in lowercase. Mutations are indicated in bold. PAM included in the target sequences.