CHALMERS
UNIVERSITY OF TECHNOLOGY

Evaluation of the on-/off-target DNA cleavage induced by Cas9 variants

Master's thesis in Biotechnology

Niklas Selfjord

Department of Biology and Biological Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2019

Masters' thesis 2019

Evaluation of the on-/off-target DNA cleavage induced by Cas9 variants

Niklas E. Selfjord

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Biology and Biological Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2019

Evaluation of the on-/off-target DNA cleavage induced by Cas9 variants

NIKLAS E. SELFJORD

Masters' thesis 2019

Department of Biology and Biological Engineering

Chalmers University of Technology

SE-412 96 Gothenburg

Sweden

Telephone + 46 (0)31-772 1000

Sponsored and perfomed in collaboration with:

AstraZeneca AB R&D

SE-431 50 Mölndal

Sweden

Telephone +46 (0)31-776 1000

[tryckeriets namn]

Gothenburg, Sweden 2019

Evaluation of the on-/off-target DNA cleavage induced by Cas9 variants

NIKLAS E. SELFJORD

Department of Biology and Biological Engineering

Chalmers University of Technology

Abstract

The CRISPR-Cas9 system of gene editing has future potential for therapeutic applications in humans. Exploration of novel Cas9 variants is currently expanding and improving the capabilities of genome engineering. A major obstacle to the clinical translation of Cas9 gene therapy is unintended mutations at off-target sites. The specificity of guide RNAs and Cas9 variants needs to be carefully evaluated prior to their therapeutic use. This thesis work aimed to explore on/off-target activity of Cas9 variants. In the first part, we have evaluated on/off-target profiles of two previously described guide RNAs known to be specific and promiscuous in single cells by sequencing the off-target sites in Cas9 treated mouse embryos. We have found that highly efficient editing with a specific guide can be achieved in embryos with no detectable off-targets. In addition, we have shown that highly efficient editing by a promiscuous guide can induce off-targets in embryos, and off-target profiles of individual embryos can be decidedly distinct. In the second part, we have evaluated the on/off-target activity of a novel Cas9 variant, AzCas9. We have demonstrated that AzCas9 can efficiently cleave the intended target site in mouse cells, and are currently evaluating its specificity in the mouse genome with a promiscuous guide using 'circularization for *in vitro* reporting of cleavage effects by sequencing' (CIRCLE-seq).

Keywords: CRISPR, Cas9, gene editing, cell transfection, off-target detection, CIRCLE-seq, next-generation sequencing.

Acknowledgements

I thank the whole of AstraZeneca for giving students the possibility of performing work and gaining valuable knowledge and experience within the pharmaceutical industry during their education.

I thank all members of the Precise Genome Editing team at the department of Translational Genomics for all their advice and encouragement. I give special mention to Nina Akrap, Burcu Bestas, Carl Möller, Amir Taheri-Ghahfarokhi and Sandra Wimberger. Their assistance and scientific expertise have been invaluable to me and to this project.

I also thank the NGS / Transcriptomics team at the department of Translational Genomics; Graham Belfield, Maryam Clausen, Inken Dillmann, Damla Etal, Julia Lindgren, Björn Magnusson and Anja Mezger. Their monthly coordinated sequencing work has directly contributed to this project.

I thank Mike Firth at the Quantitative Biology department for his work with sequencing-data analysis and bioinformatics.

I thank Marcello Maresca and Mohammad Bohlooly-Yaganeh at the department of Translational Genomics for their funding and management of the project.

And finally, my deepest gratitude I give to my supervisor and mentor Pinar Akçakaya. Her guidance, inspiration and good cheer has made this year the most exciting and rewarding period of all my time as a student.

Niklas Selfjord, Gothenburg, June 2019

Table of Contents

1 Introduction

Clustered regularly interspaced palindromic repeats (CRISPR) is a groundbreaking new technology that has revolutionized the field of biology[1]. CRISPR was first discovered as an adaptive immune system in bacteria and archaea, and is comprised of a number of CRISPR associated (Cas) proteins working in unison to defend against viral infections[2]. Endogenous Cas proteins capture foreign viral DNA, and uses it to integrate a novel recognition element, called a spacer, into the bacterial host genome. The spacer is then transcribed as RNA, processed by Cas proteins, and finally used as a homing device to find and destroy DNA originating from the captured virus.

Cas9 is an endonuclease that uses the processed CRISPR RNA to find and cleave double-stranded DNA, endogenous to many bacterial species[1,3]. Within the processed RNA is a guide sequence, which forms base pairs with the target DNA. Cas9 uses the RNA guide to scan any available DNA for a match. Sufficient complementarity between the guide RNA and the target DNA activates the catalytic domains of Cas9, cleaving the target DNA on both strands at the site of RNA recognition.

The potential of the RNA-guided nucleolytic activity of Cas9 was recognized early, and endogenous variants of the nuclease from bacterial species, such as *S. pyogenes* (SpCas9), were hijacked and engineered to require only two components: The Cas9 protein itself, and a short single guide RNA (gRNA), consisting of a scaffold for binding to Cas9 and a 20 nt guide sequence[1,4]. This created a simple and tremendously versatile system, where Cas9 can easily be programmed to induce DNA cleavage at a desired target site. The technology opened the doors for site-specific genome modifications; CRISPR-Cas9 has since been used to target, edit and regulate segments of the genomic DNA in a multitude of cells and organisms.

CRISPR-Cas9 technology may one day be used for therapeutic genome editing[1,5]. Many human diseases are caused by simple, heritable mutations in the genome. Inactivation of disease-causing genes by frameshift mutations, and correcting disease-causing mutations via precise knock-in of a non-mutated fragment can both be achieved by Cas9 in human cells.

One major concern with therapeutic genome editing is safety[4]. A considerable risk with CRISPR-Cas9 technology is that the guiding specificity is less than perfect. Cas9 has been shown to cleave DNA at sites with non-perfect complementarity to the guide[5,6]. Such sites are called off-targets, and presents a major obstacle to the clinical application of CRISPR-Cas9. Unwanted cleavage at off-target sites can cause knock-out mutations in functional genes, or genetic rearrangements causing genomic instability. Carefully assessing and improving the on/off-target specificity of gRNAs and Cas9 variants is crucial before therapeutic use is possible.

The Precise Genome Editing (PGE) team of AstraZeneca has identified several novel Cas9 variants. One variant, henceforth referred to as AzCas9, shows potential for future applications. The nucleolytic activity of AzCas9 has been verified in human cells, and the resulting genomic alterations could potentially be used to facilitate efficient, precise gene knock-in. The AzCas9 variant needs to be studied and optimized prior to further applications. An essential part of the screening and optimization process of AzCas9 is to evaluate the editing efficiency and specificity of its on/off-target nuclease activity, both *in vitro* and in cells, as well as evaluating its capability for efficient knock-in.

1.1 Aim

The aim of this project is to evaluate the on/off-target nuclease activity of Cas9 variants.

1.2 Objectives

To achieve the stated aim, three objectives were established.

1) Investigating SpCas9-induced off-targets in mouse embryos

In a study by Akcakaya et al., a strategy was developed, called 'verification of in vivo off-targets' (VIVO), to identify and evaluate the off-target effects of gene-editing nucleases *in vivo*[5]. In VIVO, a superset of potential off-target sites is first identified *in vitro* using the off-target detection method 'circularization for *in vitro* reporting of cleavage effects by sequencing' (CIRCLE-seq). Sites identified are subsequently evaluated *in vivo* for indel mutations by targeted amplicon sequencing using genomic DNA extracted from livers of nuclease-treated animals. The strategy was used to assess off-target effects in mice infected with an adenoviral vector encoding SpCas9 and gRNA. A 'promiscuous' gRNA (gP), deliberately designed to have a high likelihood of inducing off-target mutations in the mouse genome was tested, as well as two gRNAs designed to have higher orthogonality (gM and gMH), all targeting the mouse *Pcsk9* gene. The study found that by appropriate design of gRNAs, efficient *in vivo* on-target editing can be achieved with no detectable off-target mutations.

Even though no off-target mutations were detected, undesired editing can still occur at frequencies lower than the current limit of targeted amplicon sequencing detection. A further development of the study is to investigate whether enriching the editing efficiency with the same gRNAs reveals low frequency mutations at off-target sites. One plausible way of increasing the editing efficiency as compared to adenovirus infection in adult mice is to directly inject mouse embryos with Cas9 mRNA and gRNA. Cas9 editing occurring in one or two-cell stage embryonal cells will be inherited by all subsequent generations of cells and should thus result in higher editing rates overall. Results will also provide information on individual off-target profiles in single cells.

To evaluate if enriching the Cas9 on-target editing efficiency results in significant off-target mutations, targeted amplicon sequencing was performed on genomic DNA extracted from embryos injected with mRNA coding SpCas9 and either of the gP and gMH gRNAs. On- and off-target sites for both gRNAs, previously identified by CIRCLE-seq and verified in mouse livers, were amplified and sequenced by next-generation sequencing (NGS). The resulting data was used to assess the editing specificity of SpCas9 with gP and gMH in mouse embryos.

2) Evaluation of on/off-target editing outcomes by novel Cas9 variants in mouse cells

The on-target editing efficiency of AzCas9 was tested in mouse cells. Mouse cells were treated with AzCas9 and gRNA through plasmid transfection. Two gRNAs, Nfx and gP (targeting the mouse *Nfx1* and *Pcsk9* genes, respectively), were tested for each variant. Plasmids expressing the guides with compatible RNA scaffolds for SpCas9 and AzCas9 were cloned. Plasmids expressing both guides with a second version of the AzCas9 scaffold, found to more efficient in human cells, was cloned and tested in mouse cells. Editing outcomes were evaluated by NGS and compared with the widely used SpCas9 variant.

In certain experiments short, double-stranded DNA donors with varied structures were co-transfected, to evaluate the efficacy of knock-in after Cas9 variant-induced cleavage. Work on knock-in experiments are currently ongoing.

Since Cas9 nucleases can cause unwanted mutations at sites with sequence similarity to the desired target, optimizing nuclease and gRNA specificity is crucial before any future clinical applications. The on/off-target specificity of AzCas9 with the promiscuous gRNA gP was evaluated using CIRCLE-seq. CIRCLE-seq provides a list of *in vitro* off-target sites, ranked by

read-count. Identified off-target sites will be investigated by targeted amplicon sequencing on genomic DNA extracted from nuclease-treated mouse cells.

3) Cloning AzCas9 under an inducible promoter

A commonly used method of delivery for Cas9 to cells is through transient transfection of plasmids coding for the protein and gRNAs. This method is dependent on efficient transfection, which can be difficult to achieve in many cell lines because of the size of Cas9 proteins. An alternative method of Cas9 delivery would be to generate stable cell lines with inducible expression of the protein. Generation of cell lines with inducible expression of AzCas9 would be convenient in experiments requiring higher throughput, such as testing gRNA scaffolds and nuclease activity in the presence of drug molecules. Temporally controlled expression of AzCas9 would be guaranteed in all cells used, and only the gRNAs would need to be transfected to start nucleolytic activity.

Generation of stable cell lines with inducible expression of AzCas9 can be achieved through 'Obligate Ligation-Gated Recombination' (ObLiGaRe)[7]. ObLiGaRe uses a pair of zinc finger nucleases (ZFN) to precisely integrate a section of a donor vector into the genome of a target organism. By using ZFNs targeting genomic safe harbors like the *Rosa26* and *AAVS1* loci, an engineered gene construct can be safely integrated into mouse and human cell lines without disrupting normal gene function[8].

An ObLiGaRe donor vector containing AzCas9 under an inducible promotor was designed, which would enable temporal expression of AzCas9 in cells by addition of doxycycline. The cloning of this vector was set as an optional objective for the project. To achieve this, an AzCas9 cassette should be cloned into an ObLiGaRe donor vector backbone, previously designed and manufactured by the PGE team, through Gibson assembly. The ObLiGaRe donor backbone contains the inducible promoter and selection markers, which can be used to generate cell lines with stable integration of the construct. The vector will be used to transfect mouse and human cell lines together with a vector expressing a pair of ZFN, which enables insertion of the inducible AzCas9 into the *Rosa26* and *AAVS1* loci in mouse and human genomes, respectively.

2 Theory

2.1 CRISPR-Cas9

CRISPR-Cas9 nucleases are a class of proteins that have transformed the capabilities of genome editing in a multitude of cells and organisms[1,4]. Cas9 nucleases are RNA-guided endonucleases that can be programmed to efficiently cleave double stranded DNA at the target site. The most widely used orthologue of Cas9 originates from the bacterial species *Streptococcus pyogenes* (SpCas9) and has been engineered to require only two components: the Cas9 protein itself, and an approximately 100 nt long guide RNA (gRNA).

The mechanism of Cas9 induced cleavage is shown in *Figure 2.1*. The 20 nt long 5' end of the gRNA (crRNA) is complementary to the target DNA that needs to be followed by a protospacer adjacent motif (PAM) sequence, which in the case of SpCas9 is the three-nucleotide 5'-NGG-3' motif. The rest of the gRNA (tracrRNA) serves as a scaffold for complexing with the Cas9 protein. The Cas9 protein recognizes the specific PAM sequence and initiates unwinding of the neighboring DNA duplex, allowing the crRNA to base-pair with the target sequence. Adequate RNA-DNA complementarity induces a conformational change in the Cas9 protein, activating the two nuclease domains which results in a blunt-end double stranded break (DSB).

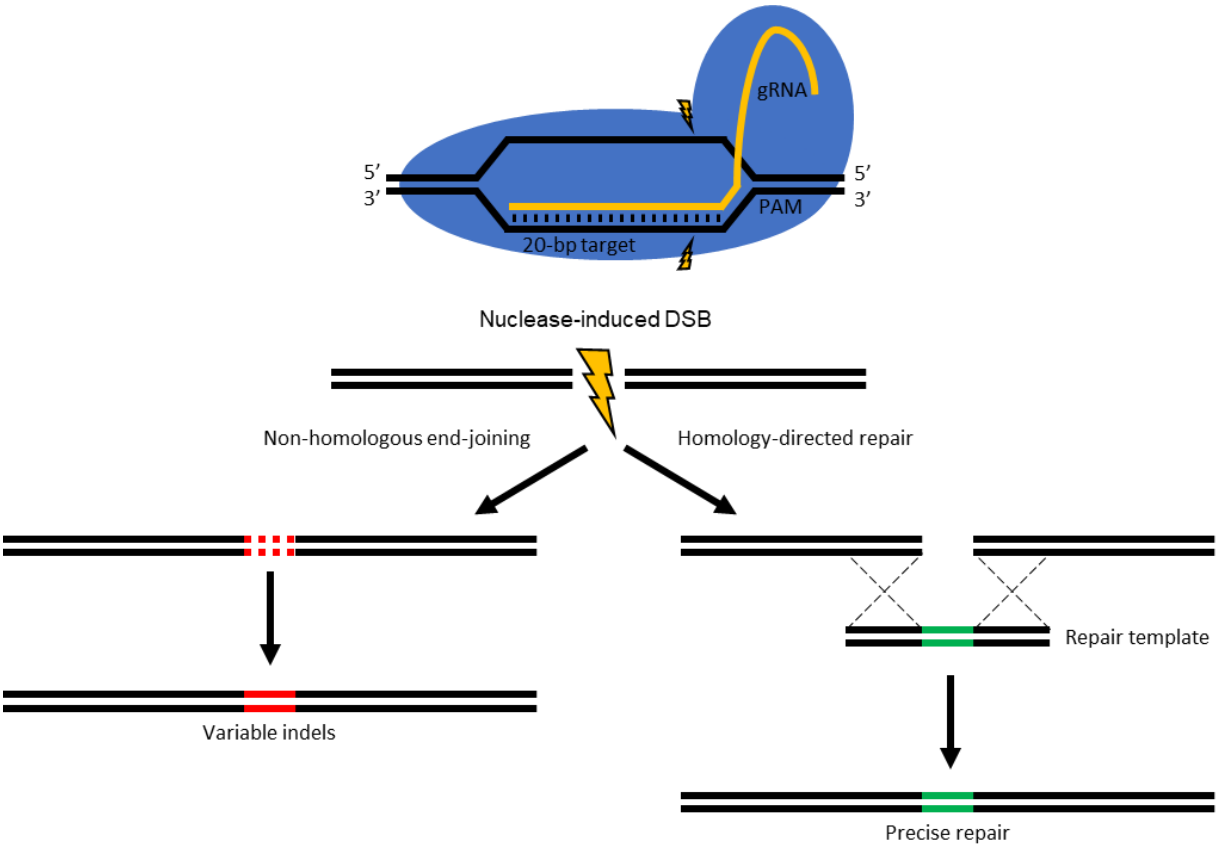

Figure 2.1. Mechanism and main repair outcomes of CRISPR-Cas9 induced DNA cleavage.

DSBs in cells can then be repaired by one of two main DNA damage repair mechanisms: non-homologous end-joining (NHEJ), which is error-prone and often results in frame-shifting insertion or deletion (indel) mutations causing a knock-out, or the less-frequent but accurate homologous recombination (HR) pathway, which uses a homologous donor template to precisely repair the damage (*Figure 2.1*).

CRISPR-mediated gene editing has future potential for clinical applications, e.g. inactivation of disease-causing genes via knock-out or correcting disease-causing mutations via precise gene knock-in. A major obstacle before any such application is the inherent off-target activity of Cas9 nucleases. Cas9 has been shown to cleave DNA at sites with up to 6 mismatches between the target and gRNA sequence[4-6]. Undesired cleavage at off-target sites can cause disruption of normal gene function. Carefully assessing and improving the on/off-target specificity of gRNAs and Cas9 variants is crucial before therapeutic use is possible.

2.2 Determining the specificity of CRIPSR-Cas9

Early studies showed that Cas9 cleavage can cause mutagenesis at mismatched sites at similar or even higher frequencies than the intended target[6,9]. Based on these and other results, *in silico* prediction tools such as E-CRISP were developed to predict off-target sites based on sequence similarity[10]. *In silico* prediction followed by targeted sequencing provided an initial method for off-target identification. One major drawback with this method is that predictions are not made in an unbiased manner, meaning that sites not fitting the criteria of the prediction algorithm might never be discovered or evaluated.

To date, several experimental strategies have been published to evaluate the specificities of CRISPR-Cas9 systems in an unbiased and genome-wide manner, each with its own strengths and drawbacks[4]. These strategies can be divided into cell-based, *in situ* and *in vitro* assays.

Cell-based assays provide rapid detection of off-targets in environments approximating *in vivo*, but detect off-targets after repair outcomes. Since a portion of Cas9-induced DSBs are precisely repaired without any mutation, some off-target sites can be missed. *In situ* assays can be used to detect transient DSBs in fixed cells before repair, but not sites that are cleaved and repaired before or after the permeabilization occurs. *In vitro* assays using extracted DNA provide a valuable complement to cell-based and *in situ* assays. The ability to tightly regulate purified Cas9-gRNA complex concentrations and treatment times, as well as bypassing cell-based factors (e.g. chromatin context, localization) enables maximum off-target cleavage, leading to more comprehensive detection and more reproducible results. However, sites that are cleaved *in vitro* are not guaranteed to be cleaved in cells or *in vivo*, and would need verification in such a context before being considered a bone fide off-target site. A summary of current methods for CRISPR-Cas9 off-target detection is shown in *Table 2.1*.

Table 2.1. Summary of current methods for CRISPR-Cas9 off-target detection.

	Detection using			Detection stage					
Method	**DNA**	**Cells**	**Tissues**	**No repair**	**Pre-repair**	**Post-repair**	**Sensitivity**	**Specificity**	**Reference**
Digenome-seq/ DIG-seq	+			+			+++	+	Kim, 2013/ Kim, 2018
CIRCLE-seq	+			+			+++	+	Tsai, 2017
SITE-seq	+			+			+++	+	Cameron, 2017
GUIDE-seq		+				+	++	++	Tsai, 2016
IDLV capture		+				+	++	++	Wang, 2015
HTGTS/ LAM-HTGTS		+				+	+	++	Frock,2015/ Hu, 2016
BLESS/ End-seq/ DSB-capture		+			+		++	++	Crosetto,2013/ Canela, 2016/ Lensing, 2016
BLISS		+	+		+		++	++	Yan, 2017
DISCOVER-seq		+	+		+		+	+++	Wienert, 2019

'Genome-wide unbiased identification of DSBs enabled by sequencing' (GUIDE-seq) is a cell-based assay, which is based on the integration of a double-stranded oligodeoxynucleotide (dsODN) tags at Cas9 induced DSBs, followed by tag-specific amplification and deep sequencing (*Figure 2.2.a*)[11]. GUIDE-seq is highly sensitive and experimentally simple but requires cells that can be efficiently transfected with the short dsODN tags. GUIDE-seq is also dependent on the endogenous NHEJ repair pathway, which can cause cell-line specific differences in results depending on the repair pathway preference of the cell-type examined.

'Breaks labelling, enrichment on streptavidin and next-generation sequencing' (BLESS) is an *in situ* assay, used for genome-wide detection of DSBs in fixed cells (*Figure 2.2.b*)[12]. Nuclei of nuclease-treated cells are fixed and permeabilized, and transient DSBs are detected by *in situ* ligation of biotinylated hairpin adaptors. BLESS can be applied to examine tissues treated with Cas9 *in vivo,* but can only detect sites that have already been cleaved at the moment of permeabilization, requires extensive amounts of input material and can be challenging to perform experimentally.

'Breaks labeling *in situ* and sequencing' (BLISS) is a more recent *in situ* assay designed to circumvent some of the limitations of BLESS[13]. In BLISS, Cas9-gRNA treated cells or tissue sections are fixed onto a microscope slide, and transient DSBs are *in situ* blunted. dsDNA adapters containing a T7 promoter are ligated to blunted ends, and the DNA is purified and

fragmented by sonication. DNA fragments with ligated adapters can then be linearly amplified by *in vitro* transcription and sequenced by NGS. BLISS provides a more sensitive and less laborious *in situ* alternative to BLESS.

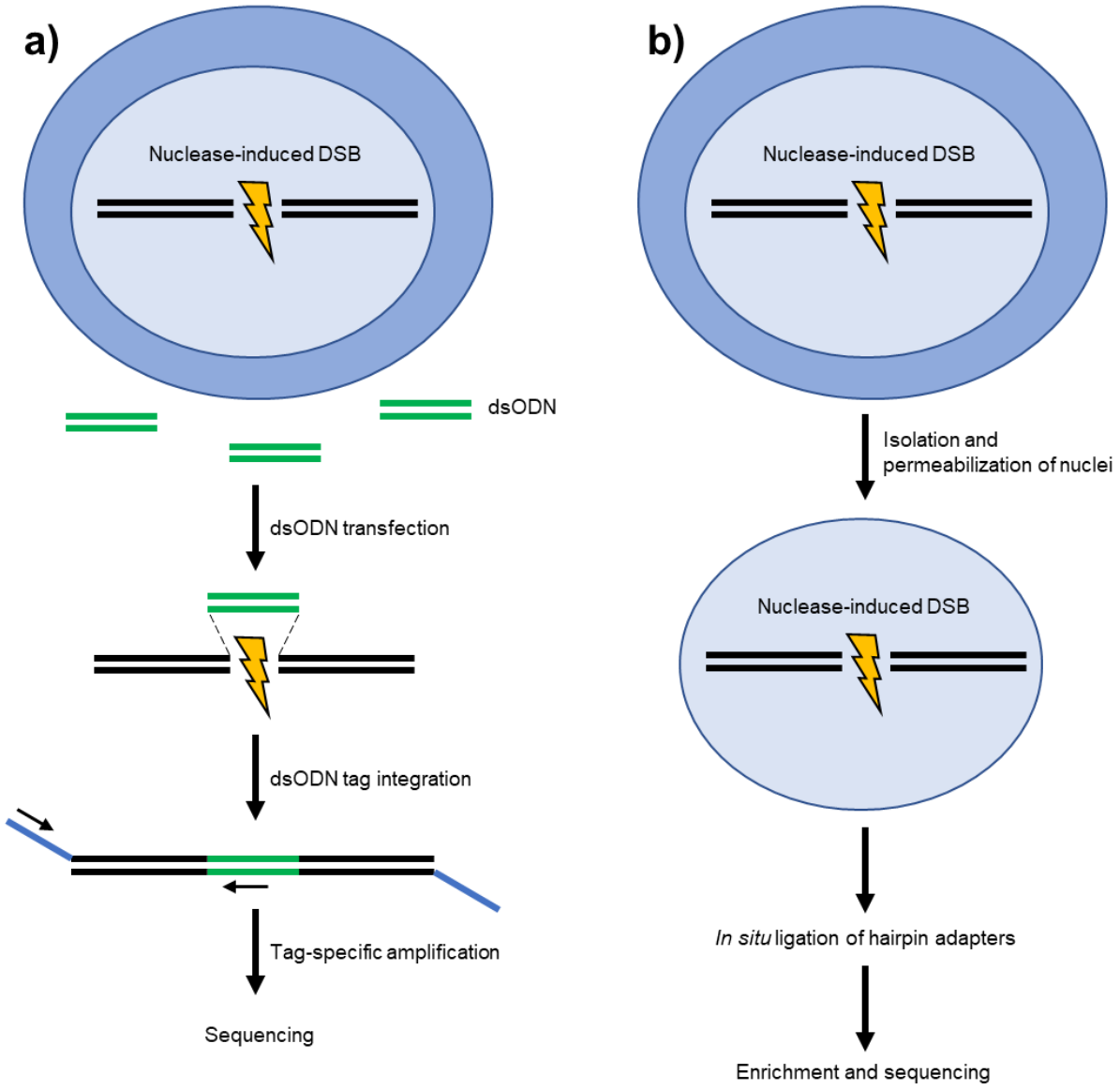

Figure 2.2. Examples of cell-based and *in situ* genome-wide off-target detection methods. a) 'Genome-wide unbiased identification of DSBs enabled by sequencing' (GUIDE-seq). b) 'Breaks labelling, enrichment on streptavidin and next-generation sequencing' (BLESS).

'Digested genome sequencing' (DIG-seq) is an *in vitro* assay, where extracted genomic DNA is digested by Cas9-gRNA and then sequenced to very high coverage with NGS[14]. Enrichment of reads with the same start and end position are used to identify Cas9-induced DSBs. Because cleaved sequences are not amplified in any way, DIG-seq is highly sequencing-inefficient. Most of the reads are spent on background sequences not related to nuclease-induced DSBs, therefore the method requires around 400 million reads per sample and the background levels are high.

'Circularization for in vitro reporting of cleavage effects by sequencing' (CIRCLE-seq) is a more recently described *in vitro* assay, providing a solution to the disadvantages of DIG-seq[15]. In CIRCLE-seq, genomic DNA is first sheared and circularized by intramolecular ligation. Circular DNA is purified, and treated with Cas9-gRNA. Any sequence containing an on- or off-target

site will be cleaved and linearized, leaving open ends for adaptor ligation, amplification and deep sequencing. CIRCLE-seq provides a sensitive and sequencing-efficient way of assaying Cas9 off-target sites *in vitro*, but still suffers from a high false discovery rate.

'Discovery of *in situ* Cas off-targets and verification by sequencing' (DISCOVER-seq) is a novel *in situ* method recently published, that utilizes chromatin immunoprecipitation to detect CRISPR-Cas off-targets[16]. After Cas9-gRNA treatment in cells or *in vivo*, endogenous DNA repair factors localize to the site of the DSB. In DISCOVER-seq, a DNA repair factor, e.g. MRE11, is first crosslinked to the ends of Cas9-induced DSBs. DNA-protein complexes are then sheared by sonication to a length of several hundred bp, and the repair factor along with the bound DNA is selectively immunoprecipitated using a protein-specific antibody. DNA fragments bound to the repair factor can then be purified and sequenced. Enrichment of sequencing reads at specific genomic locations are then used to determine the location and molecular characteristics (e.g. the precise location of the break on either strand) of the Cas9-induced DSB. DISCOVER-seq has a low false positive rate, but may be unable to distinguish putative off-targets detected by more sensitive methods, such as GUIDE-seq or CIRCLE-seq.

2.3 Improving the specificity of CRISPR-Cas9

Various approaches to reduce off-target effects of Cas9-mediated cleavage have been published[4]. These approaches can be categorized as either increasing the specificity of the Cas9-gRNA target cleavage or limiting the activity of the Cas9-gRNA complex to avoid accumulation of off-target mutations. Examples of improving CRISPR-Cas9 specificity are shown in *Figure 2.3*. One is via truncated guide-RNAs (tru-gRNAs), where the gRNA is shortened 2-3 bp at the 5' end to reduce tolerance for sequence mismatches as compared to full-length gRNA[17]. Another example is using paired Cas9 nickases (Cas9n), whereby a pair of mutated Cas9 proteins are directed by two gRNAs to create single-stranded breaks (nicks) on opposite DNA strands[18]. Single stranded breaks are achieved by mutating one of the two Cas9 nuclease domains to make it catalytically inactive. Cas9 proteins have also been engineered to increase on-target specificity by substitutions of amino acids involved in the non-specific interactions between the protein with the target and non-target DNA strand[19,20].

2.4 Cas protein variants

Many different orthologues and variants of Cas proteins have been described in the literature[3]. CRISPR is a bacterial adaptive immune system, evolved through competition with viruses[2]. This evolution has given rise to a diverse array of naturally occurring CRISPR-Cas systems. CRISPR-Cas systems are currently divided in two classes, with Class I consisting of systems where the effector protein (nuclease) is made up of multiple subunits, and Class II where the effector protein is a single component, e.g. Cas9.

Naturally occurring Class II Cas orthologues from many different bacterial species have been explored for genome editing (examples shown in *Table 2.2*). Cas nuclease orthologues have been shown to include variations in properties such as PAM specificity, target sequence length, gRNA length, protein size, and whether the protein targets DNA or RNA[21-24]. Some nucleases have also been found to produce "sticky" cuts, where the breaks on opposite DNA strands are several nucleotides staggered, relative to each other. Naturally occurring Cas variants have also been engineered to recognize PAM-sequences other than the 'canonical' motif, by introducing mutations in the PAM-interacting domains of the wild-type protein[25,26]. The widely used SpCas9 variant has for example been engineered to prefer an NGAG or NGCG PAM, rather than the canonical NGG motif. Cas protein variants with altered PAM specificity allow a broader range of target sites. The diversity in Cas variants and orthologues can be exploited to the find the ideal protein for specific genome engineering applications.

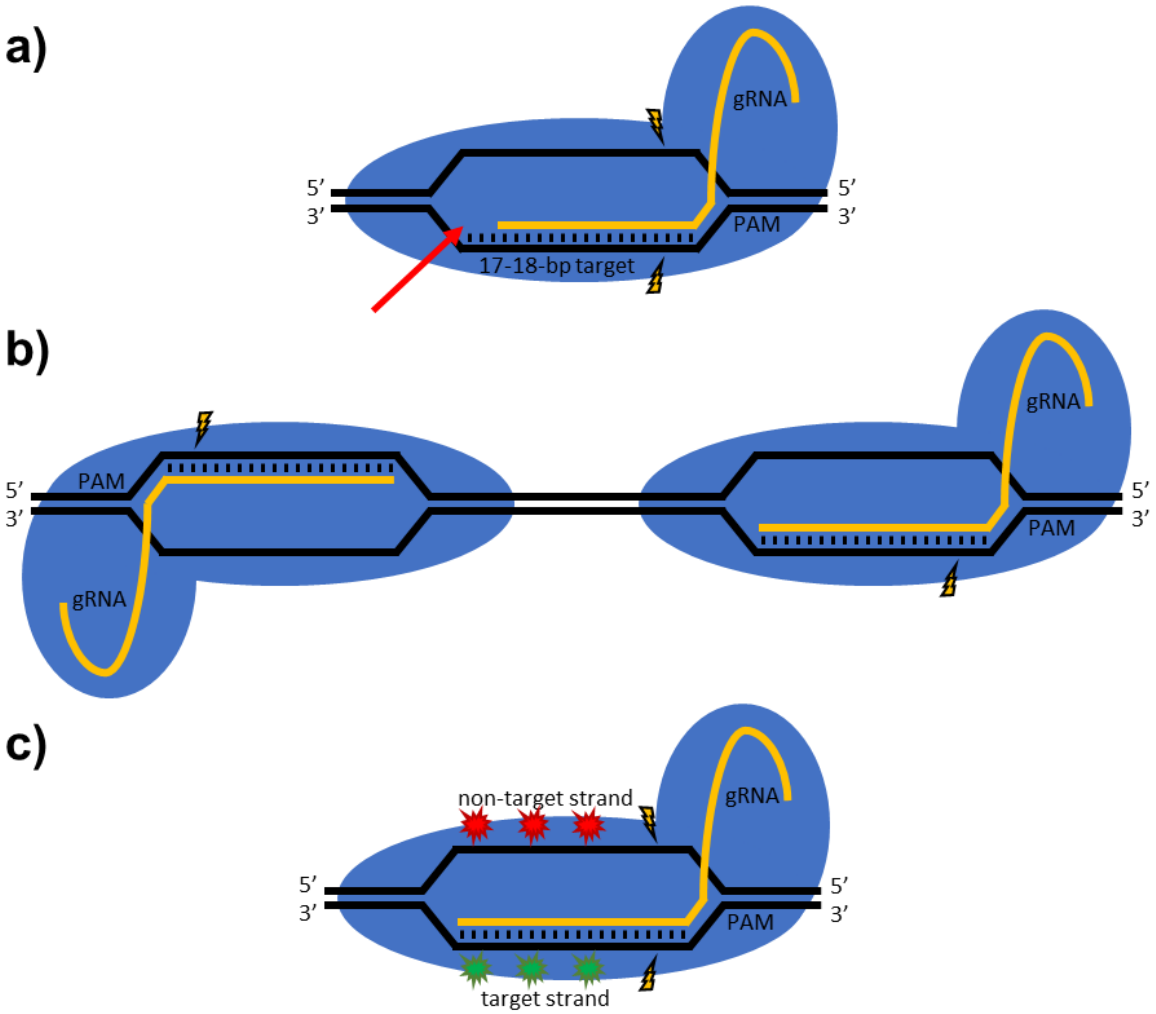

Figure 2.3. Examples of improving CRISPR-Cas9 specificity. a) Truncated gRNA (tru-gRNA); gRNA is shortened 2-3 bp at the 5'-end. b) Paired Cas9-nickases (Cas9n); Paired Cas9 with one nuclease-domain catalytically inactive make single-stranded breaks (nicks) on opposite strands, resulting in a DSB. Here shown in the 'PAM-out' orientation. c) Engineered Cas9 variants; Engineered Cas9 proteins with amino acid substitutions at residues interacting with the target (green) or non-target (red) DNA strand.

Table 2.2. Characteristics of Class II CRISPR-Cas orthologues and variants used for genome editing.

Species	Effector (variant)	Size (aa)	PAM	Target length	Comment
S. pyogenes	SpCas9	1368-1424	NGG	20 nt	Efficient, widely used
S. pyogenes	SpCas9 VQR	1372	NGAG	20 nt	Engineered PAM
S. pyogenes	SpCas9 VRER	1372	NGCG	20 nt	Engineered PAM
S. aureus	SaCas9	1053	NNGRRT	20-24 nt	Small size
N. meningitidis	NmCas9	1109	NNNNGATT	23, 24 nt	Specific PAM
F. novicida	FnCas9	1629	NGG	22 nt	Sticky cut
L. bacterium	Cas12a	1228	TTTV	23, 24 nt	No tracr, sticky cut
L. buccalis	Cas13a	1159	'3 A, U, or C	N/A	No tracr, RNA target

3 Methods

3.1 Delivery of Cas9 and gRNA to cells

The process of delivering foreign nucleic acids into eukaryotic cells is called transfection. There are many different methods of delivering DNA to mammalian cells cultured *in vitro*[27]. Foreign DNA must successfully be delivered to the nucleus before it can be expressed, which includes crossing several different barriers, e.g. cell membranes and intracellular degradation

machinery. Methods of delivery include viral delivery, where DNA is encapsulated in a viral vector which is used to infect the cells, and non-viral delivery, e.g. electroporation, where pores in the cell membrane are opened using electrical currents, and complexation of DNA with cationic materials to form nanoparticles which can subsequently be endocytosed. Non-viral delivery methods have the benefit of being less immunogenic to most cell types, and being easy and safe to perform experimentally. For this project, a nanoparticle-mediated delivery method of transfection was used, using the commercially available FuGENE® HD transfection reagent. FuGENE® HD is a proprietary mixture of lipids and other components that is used to form complexes with purified DNA, in the form of plasmids and oligonucleotides, that can then fuse with the cell membrane and transport its content inside. Mouse cells were co-transfected with plasmid vectors coding Cas9 protein and gRNA separately, as well as short double-stranded oligonucleotide (dsODN) donors for evaluation of gene knock-in.

3.1.1 Cas9 delivery vector

A schematic of the plasmid vector used to deliver Cas9 expression in mouse cells is shown in *Figure 3.1*. The Cas9 delivery vector contains a Cas9-T2A-GFP cassette under control of a CMV promoter, an f1 origin of replication (ORI), a neomycin resistance (NeoR) selection marker under control of SV40 early promoter, a bacterial ORI and an ampicillin resistance (AmpR) selection marker. The CMV promoter is a strong constitutive RNA polymerase II (Pol II) promoter for expression of mRNAs in mammalian cells. The Cas9-T2A-GFP coding sequence codes the mRNA for the Cas9 protein, a T2A self-cleaving peptide and green fluorescent protein (GFP) in a single transcript. At the end of the Cas9 sequence is a nuclear localization signal (NLS) and a tandem 3xFLAG-tag. The NLS is used to facilitate transportation of Cas9 back into the nucleus after protein translation, and the FLAG-tags are artificial antigens that can be used in assays where antibody recognition is required, e.g. Western blotting. The T2A peptide is a short peptide linker, which is cleaved after translation. After translation of the whole Cas9-T2A-GFP fusion protein, the T2A linker is cleaved resulting in independent Cas9 and GFP monomers. Expression of GFP can be observed using fluorescence microscopy to quantify transfection efficiency and to estimate Cas9 expression. A BGH pA sequence polyadenylates the full Cas9-T2A-GFP transcript.

The f1 phage ORI enables single stranded replication and packaging into phage particles. The NeoR selection marker confers resistance to the antibiotic neomycin and can be used for selection of successfully transformed bacteria or transfected mammalian cells. The bacterial ORI is the sequence at which replication is initiated in bacteria, enables plasmid replication in *E. coli*. The AmpR selection marker confers resistance to the antibiotic ampicillin and is used for selection of successfully transformed bacteria.

3.1.2 gRNA delivery vector

A schematic of the plasmid vector used to deliver gRNA expression in mouse cells is shown in *Figure 3.2*. The gRNA delivery vector contains the single gRNA (sgRNA) cassette under a U6 promoter, a bacterial ORI and a kanamycin resistance (KanR) selection marker. The U6 promoter is a constitutive promoter for RNA polymerase III (Pol III) which transcribes DNA into 5S rRNA, tRNA and other small RNAs in mammalian cells. RNAs transcribed by Pol III do not get polyadenylated and thus remain in the nucleus. The sgRNA cassette codes the precise sequence of the gRNA (crRNA and tracrRNA) in a single transcript. The bacterial ORI is the sequence at which replication is initiated in bacteria, which enables plasmid replication in *E. coli*. The KanR selection marker confers resistance to the antibiotic kanamycin and is used for selection of successfully transformed bacteria.

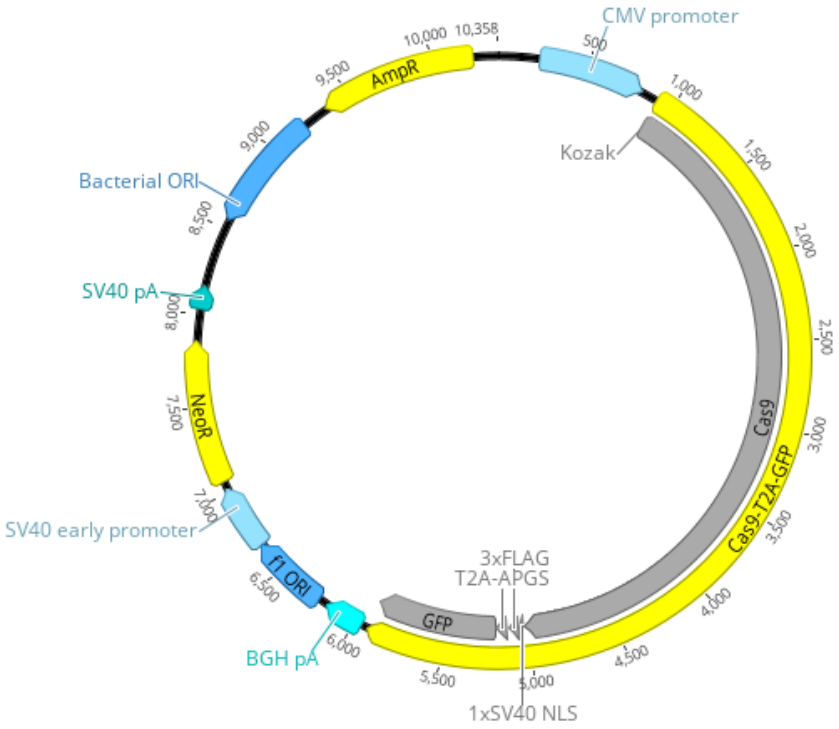

Figure 3.1. Schematic of Cas9 delivery vector. The Cas9 delivery vector contains a Cas9-T2A-GFP cassette under control of a CMV promoter, an f1 origin of replication (ORI), a neomycin resistance selection marker under control of SV40 early promoter, a bacterial origin of replication (ORI) and an ampicillin resistance (AmpR) selection marker. Image taken from Geneious version 2019.1 created by Biomatters. Available from https://www.geneious.com.

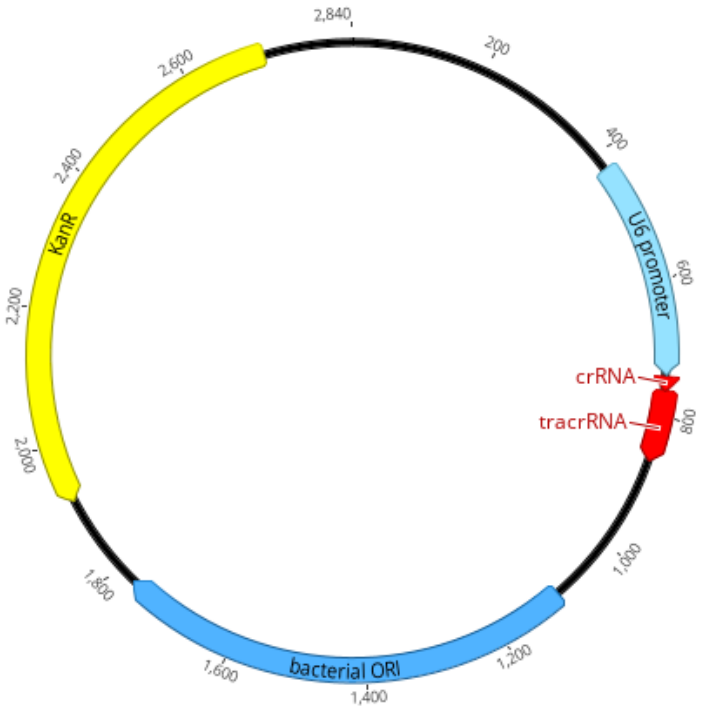

Figure 3.2. Schematic of gRNA delivery vector. The gRNA delivery vector contains the single gRNA cassette under a U6 promoter, a bacterial origin of replication (ORI) and a kanamycin resistance (KanR) selection marker. Image taken from Geneious version 2019.1 created by Biomatters. Available from https://www.geneious.com.

3.2 T7 Endonuclease I assay

The T7 Endonuclease I (T7EI) assay provides a rapid way of evaluating gene editing outcomes after nuclease treatment[28]. An overview of the T7EI assay is shown in *Figure 3.3*. The nuclease target site is first amplified from the genomic DNA of Cas9 treated samples. A portion of products will contain variable indels resulting from the imprecise repair on DSBs created by Cas9. The products are denatured into single-stranded DNA and re-hybridized into duplexes, some of which will contain mismatches at the mutated site. Samples are treated with T7EI, which recognizes and cleaves double-stranded DNA at mismatched sites with a resolution of 2 or more bp. Resulting products can then be visualized using gel electrophoresis, where the presence of shorter, mutant bands indicates Cas9 editing. One significant drawback is that T7EI assay generally underreports indel frequencies as compared with targeted deep sequencing by NGS[29], because 1 bp indels constitute a significant proportion of mutagenesis caused by Cas9 editing. Although targeted NGS is the preferred method for quantifying nuclease-mediated gene editing, the T7EI assay provides a quick and cost-effective alternative for verifying successful editing.

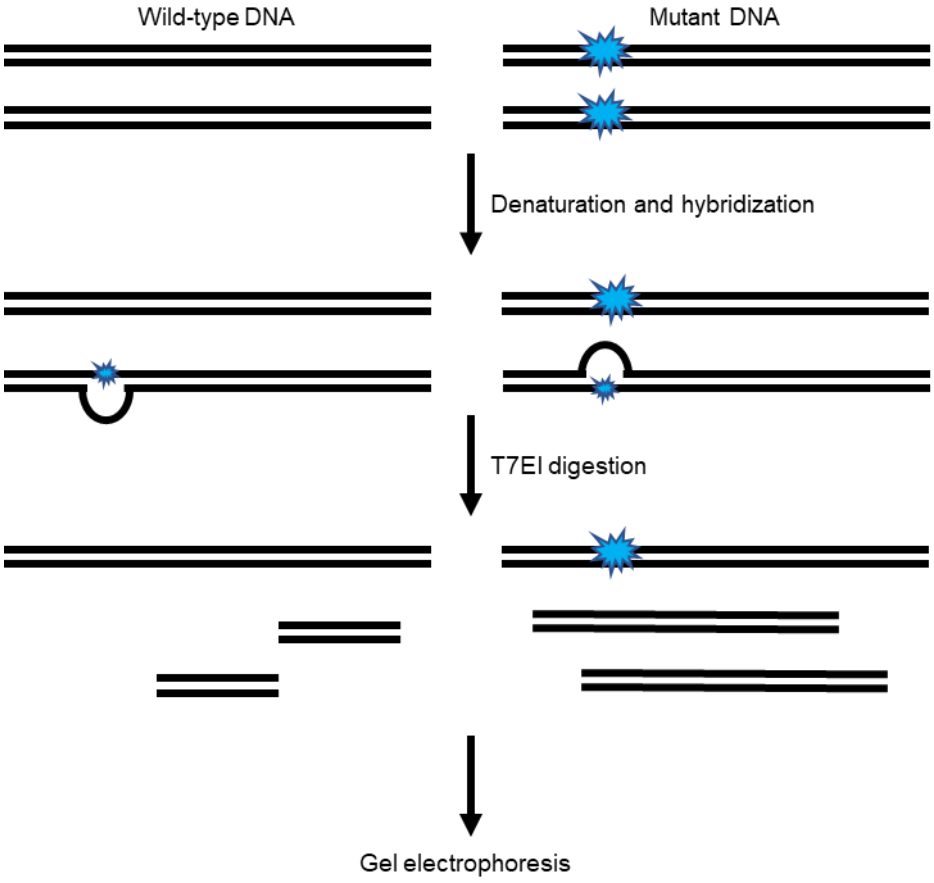

Figure 3.3. Overview of the T7 Endonuclease I assay.

3.3 CIRCLE-seq

'Circularization for *in vitro* reporting of cleavage effects by sequencing' (CIRCLE-seq) is a recently described method for identifying genome-wide off-target mutations of CRISPR-Cas9 *in vitro*[15]. The method is rapid, sensitive and accessible, providing a valuable alternative to existing cell-based methods. *In vitro* strategies for identifying off-target DSBs can have several advantages over cell-based approaches; including reproducibility, avoiding the need for efficient cellular transfection and identification of sequences rarely cleaved *in vivo*, since the concentration of active nuclease can be closely regulated. Because of its high sensitivity,

CIRCLE-seq provides a comprehensive list of potential off-target sites for a given combination of gRNA and Cas9 nuclease. The same sites can then be assessed for indels after Cas9 treatment in cells or *in vivo*[5].

The principle steps of CIRCLE-seq (*Figure 3.4*) are as follows: genomic DNA is first sheared and circularized by intramolecular ligation. Remaining linear DNA molecules are then degraded by exonuclease treatment. Circular DNA molecules are treated with Cas9 nuclease. Any circular DNA molecule containing a Cas9 cleavage site (on- or off-target) are subsequently cleaved and linearized, leaving free ends for adapter ligation, PCR amplification, and paired-end high-throughput sequencing.

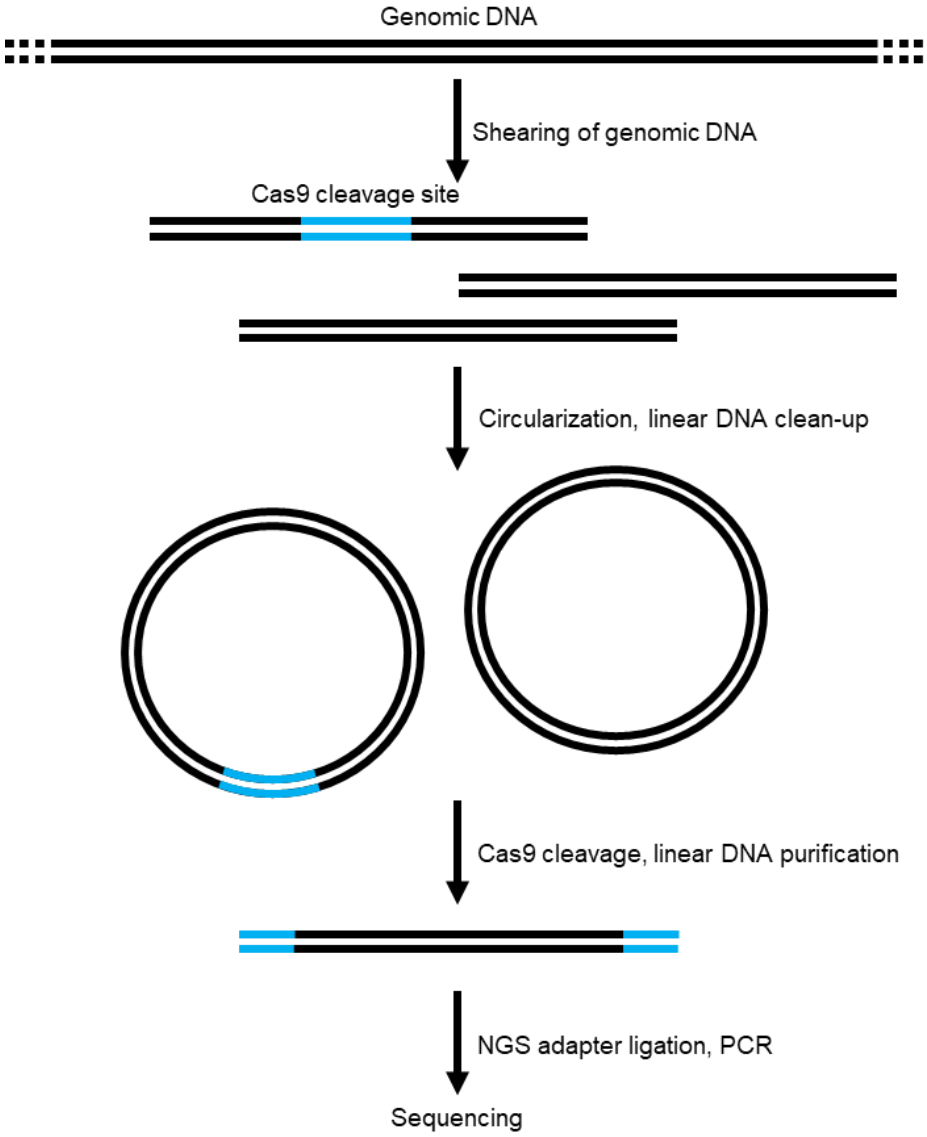

Figure 3.4. Summary of CIRCLE-seq.

3.4 Gibson assembly

Gibson assembly is a molecular cloning method of assembling two or more separate DNA fragments into a single construct[30]. Gibson assembly occurs in a single, isothermal reaction using a combination of enzymes: an exonuclease, a DNA polymerase and a DNA ligase. The principle steps of Gibson assembly are shown in *Figure 3.5.* Overlapping, homologous regions (usually 20-50 bp) are first added to fragments using PCR. Fragments and enzymes are incubated at 50 °C for 15-60 min depending on the number and size of fragments to create

3`overhangs. A 5'-exonuclease is used to chew back the 5'-ends of all fragments, overlapping 3'-ends of fragments can then anneal. Strands are extended by DNA polymerase and finally sealed by DNA ligase. Gibson assembly provides a simple and rapid means of producing potentially very large DNA constructs.

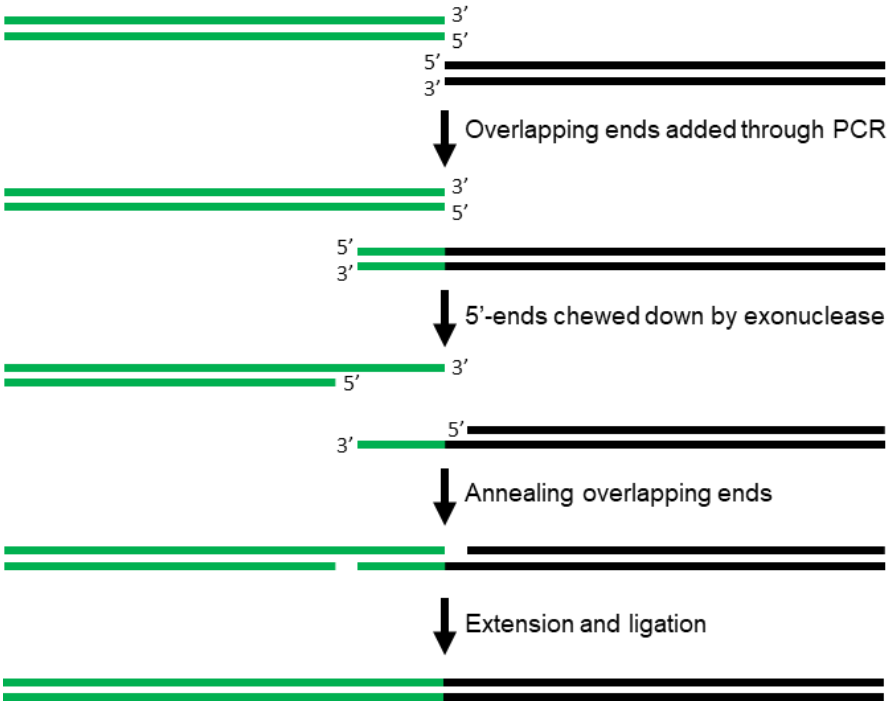

Figure 3.5. Gibson assembly of two DNA fragments.

3.5 ObLiGaRe

'Obligate Ligation-Gated Recombination' (ObLiGaRe) is a method for site-specific gene knock-in by NHEJ. The method utilizes a pair of custom-designed ZFN to make a DSB in the genome, targeting e. g. a safe-harbor locus like human *AAVS1* or mouse *Rosa26*. A circular DNA donor vector is simultaneously added, containing the same ZFN recognition site as the targeted genomic site. The donor vector is cut and linearized *in situ* by the ZFN pair, after which it can anneal to the complementary ends generated by the genomic DSB and can be integrated into the genome by the NHEJ repair pathway.

A schematic of the ObLiGaRe donor vector used for generation of cell lines with stable, inducible expression of Cas9 is shown in *Figure 3.6*. The original backbone vector (*Figure 3.6.a*) contains ZFN recognition sites for integration into *AAVS1* or *Rosa26*, the doxycycline inducible promoter element (Tet-On 3G under CAG promoter), a ccdb cassette, f1 and bacterial ORI and AmpR and NeoR selection markers. The ccdb cassette codes for a toxic protein, which effectively prevents non-resistant bacterial cells from propagating. The ccdb gene is excised from the vector before transformation and is used as a negative selection marker, ensuring that any cells surviving after transformation does not contain the original or a re-ligated ccdb intermediate vector. The AmpR and NeoR genes are used as positive selection markers for bacterial and mammalian cells, respectively.

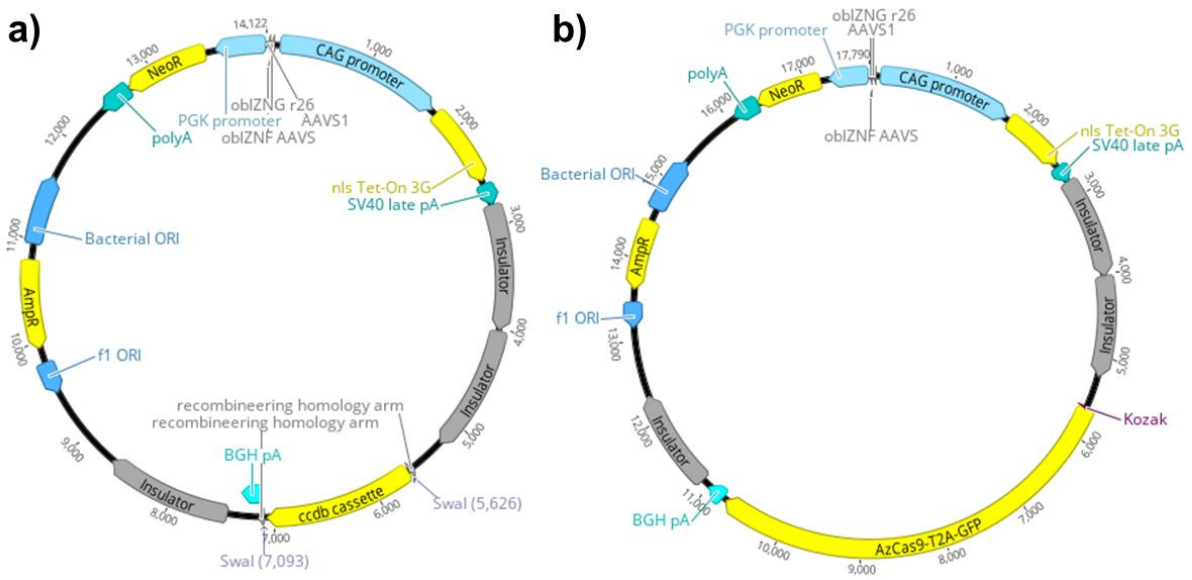

Figure 3.6. Schematic of ObLiGaRe donor vector. a) ccdb intermediate backbone vector. b) Predicted final product. The ccdb cassette is excised using restriction enzymes (SwaI), and an AzCas9-T2A-GFP cassette is inserted in its place using Gibson assembly. Image taken from Geneious version 2019.1 created by Biomatters. Available from https://www.geneious.com.

To clone the final ObLiGaRe donor vector, the ccdb cassette is first excised by digestion with the restriction enzyme SwaI. The linear fragment without the ccdb cassette is then purified. A PCR is run to amplify to the AzCas9-T2A-GFP cassette from its original AzCas9 expression vector, using primers with 30-50 bp ends that overlap with the SwaI digested backbone (marked recombineering homology arm in *Figure 3.6.a.*). The AzCas9-T2A-GFP cassette is then gel purified, and the two purified fragments are joined by Gibson assembly to form the final ObLiGaRe donor vector (*Figure 3.6.b.*).

3.6 Experimental

This section provides a summary of all experiments performed. A detailed, step-by-step protocol is provided in *Appendix A*.

3.6.1 Targeted deep sequencing

Editing outcomes at on- and off-target sites were determined by targeted deep sequencing. Target sites were amplified from 150 ng input genomic DNA using site-specific primers and Q5 High Fidelity polymerase (New England Biolabs). Primers span approximately 150-300 bp around the Cas9 cut site, and contain a 5'-sequencing adapter, compatible with Nextera XT Index Kit v2 (Illumina). A list of all sequencing primers used is provided in *Appendix B*. Amplicon PCR products were purified and size-selected using Agencourt AmPure XP beads (Beckman Coulter). Purified products were verified and quantified on Fragment Analyzer (DNF-915 kit, Advanced Analytical Technologies). Amplicons of up to five sites were normalized in concentration and pooled. Pooled amplicons were indexed with Nextera XT v2 primers by PCR using KAPA Hifi polymerase (Roche), purified with Agencourt AmPure XP beads, verified and quantified using Fragment Analyzer. Indexed amplicons were normalized to 20 nM and pooled. Final libraries were loaded onto an Illumina NextSeq500 for deep sequencing. Sequencing data was analyzed through the CRISPResso pipeline[31].

3.6.2 Translocation PCR

A PCR was run to check for translocation events between the on-target and off-target sites of the gP gRNA. The forward primer for the on-target site was used in conjunction with the reverse primer for an off-target site to check for PCR products, using Phusion Flash High-Fidelity

polymerase (ThermoFisher). PCR products were run on Fragment Analyzer using the Standard Sensitivity NGS Fragment Analysis kit (DNF-473).

3.6.3 sgRNA cloning

A duplexed pair of oligos coding the crRNA were cloned into a BsaI (New England Biolabs) digested base vector containing the Cas9 scaffold. Cloned sgRNA plasmids were transformed into DH10-beta (New England Biolabs) competent bacteria and amplified in 4 ml overnight cultures. A 500 µl aliquot of overnight culture was used to inoculate a 200 ml culture, and the rest was used to purify plasmids with QIAprep Spin Miniprep Kit (Qiagen). Concentration of purified plasmids were measured by Nanodrop and correct insertion was verified by Sanger sequencing. The 200 ml culture was grown overnight and verified sgRNA plasmids were purified with EndoFree Plasmid Maxi Kit (Qiagen) according to the manufacturer's instructions.

3.6.4 Mouse cell culture and transfection

Neuro2a, NIH/3T3 and primary MEF cells were obtained from the AstraZeneca Cell Bank and cultured in DMEM (Life Technologies) supplemented with 10% FBS, GlutaMax (Life Technologies) and penicillin/streptomycin at 37 °C with 5% CO_2. Cells were seeded in plates (12-well plates for Neuro2a, 24-well plates for NIH3T3 and MEF cells) a day before transfection. Neuro2a cells were co-transfected with 500 ng of Cas9 plasmid, 500 ng of sgRNA plasmid and 13 pmol dsODN at a 4.5:1 ratio of FuGENE HD (Promega) to DNA. A list of all dsODN sequences used is provided in *Appendix B*. For negative control experiments, cells were transfected with 1 µg of pmaxGFP plasmid (Lonza), 1 µg of Cas9 plasmid or FuGENE HD reagent only. For experiments in 24-well plates, all DNA amounts were halved. Genomic DNA was extracted ~72 h post-transfection with Gentra Puregene Cell kit (Qiagen) (12-well plates) or QuickExtract DNA extraction solution (Lucigen) (24-well plates) according to the manufacturer's instructions. Gene editing outcomes were evaluated by targeted deep sequencing and/or T7EI mismatch cleavage assay (Alt-R Genome Editing Detection Kit, IDT) according to the manufacturer's instructions.

3.6.5 In vitro transcription of gRNAs for CIRCLE-seq

gBlocks containing T7-promoter, crRNA and variant-specific tracrRNA were ordered from Geneart Strings DNA Fragments. All sequences are provided in *Appendix B*. gRNA gBlocks were amplified and a 3'-adenosine was added by PCR from 5 ng input DNA using AmpliTaq Gold DNA polymerase (ThermoFisher). PCR products were cloned into a TOPO TA vector (ThermoFisher), transformed into DH10-beta (New England Biolabs) competent bacteria and amplified in 4 ml overnight cultures. Plasmids were purified with QIAprep Spin Miniprep Kit (Qiagen). Purified plasmids were measured by Nanodrop and correct insertion was verified by Sanger sequencing. 10 µg verified plasmid was digested with EcoRI (New England Biolabs) and excised template DNA was extracted from agarose gel using QIAquick Gel Extraction kit (Qiagen). Template DNA was further cleaned-up using phenol-chloroform extraction. sgRNA was transcribed from template DNA using MEGAscript T7 Transcription kit and purified using MEGAclear Transcription Clean-Up kit (both from ThermoFisher).

3.6.6 CIRCLE-seq library preparation

Purified genomic DNA (C57BL/6 DNA purified with Gentra PureGene Tissue Kit, Qiagen) was sheared to an average length of 300 bp with a Covaris S2 instrument. Sheared DNA was end-repaired, A-tailed and ligated to a stem-loop adapter (sequence available in *Appendix B*) using KAPA High-throughput library preparation kit (KAPA Biosystems). Adapter-ligated DNA was treated with Lambda Exonuclease and Exonuclease I (*E. coli*), followed by treatment with USER enzyme and T4 polynucleotide kinase (all enzymes from New England Biolabs). 500 ng input DNA was circularized using T4 DNA ligase. Linear DNA molecules were subsequently removed by treatment with Plasmid-Safe ATP-dependent Dnase (Epicentre). *In vitro* cleavage reactions were performed with 90 nM Cas9 nuclease (*S. pyogenes* from New England Biolabs,

AzCas9 expressed and purified in-house), 90 nM *in vitro* transcribed gRNA and 250 ng circularized DNA. Digested, linear DNA was A-tailed, ligated with hairpin adapter and treated with USER enzyme. Resulting products were amplified and indexed by PCR with KAPA Hifi HotStart ready mix (Kapa Biosystems). PCR products were pooled and sequenced with 150 bp paired-end reads on an Illumina NextSeq500. Data was processed using the CIRCLE-seq analysis pipeline[15].

3.6.7 Inducible AzCas9 ObLiGaRe donor

ObLiGaRe donor vector containing doxycycline-inducible AzCas9 was prepared by cloning an AzCas9-T2A-GFP cassette into a SwaI (New England Biolabs) digested base vector containing the inducible promoter and neomycin resistance selection marker. The AzCas9-T2A-GFP cassette was amplified by PCR with Phusion Flash High-Fidelity polymerase (ThermoFisher), using primers that contained 30/50 bp ends homologous to the ends of the SwaI digested base vector. PCR products and digested base vectors were verified and separated on agarose gel and purified using QIAEX II gel extraction kit (Qiagen) according to the manufacturer's instructions. Purified fragments were joined by Gibson assembly Master Mix (New England Biolabs) at a vector:insert ratio of 1:4. 2 µl of the assembly reaction was used to transform DH10-beta (New England Biolabs) competent bacteria.

4 Results

4.1 Off-target editing in embryos

Indel frequencies in mouse embryos injected with mRNA coding for Cas9 and specific gMH gRNA, as measured by targeted amplicon sequencing and analyzed by CRISPResso, are shown in *Figure 4.1*. In total, the on-target site and 10 highly ranked (as determined by CIRCLE-seq[5]) were sequenced, using genomic DNA extracted from four control embryos treated with Cas9 only and four embryos treated with both Cas9 and gRNA. Off-target sites were selected for sequencing based on the CIRCLE-seq read counts and character of mismatches relative to the on-target site, with low numbers of mismatches and mismatches at the PAM-distal end of the gRNA being favored since these sites are more likely to show off-target activity[4].

On-target efficiency (site 2, *Pcsk9*) was high for all 5 gMH treated embryos, up to 89.2 % and 95.2 % editing embryos 2 and 3, respectively, indicating successful editing at a very early cellular stage in the embryo. No significant off-target activity could be detected for any of the off-target sites. Approximately 66 % indels were measured at site 6 (*Apopt1*) for all embryos, but since there was no significant difference between control and gRNA treated samples, this variation was not attributed to Cas9 nuclease activity. These results confirm gMH to be a highly specific gRNA.

Indel frequencies in mouse embryos injected with mRNA coding for Cas9 and promiscuous gP gRNA, as measured by targeted amplicon sequencing and analyzed by CRISPResso, are shown in *Figure 4.2*. In total, the *Pcsk9* on-target site and 8 previously validated gP off-target sites[5] in mouse livers were sequenced, using genomic DNA extracted from four control embryos treated with Cas9 only and four embryos treated with both Cas9 and gRNA. One site (site 8, *Marcks*) had very low read coverage, and was excluded from analysis. On-target efficiency (site 1, *Pcsk9*) was high for all four gP treated embryos, up to 99.4 % editing. Off-target editing was detected at three of the highest ranked off-target sites in two of the gP-treated embryos (embryos 2 and 3). The other two gP-treated embryos showed no detectable off-targets (embryos 1 and 4). Notably, embryos 1 and 4 exhibited lower on-target editing efficacy (33.3% and 44.8%, respectively) than embryos 2 and 3 (99.4% and 74%, respectively), and they showed higher on-target editing rate than previously reported mouse livers (up to 30.4%) [5].

Embryos 2 and 3 showed editing significantly higher than controls in site 2 (*Vps25-Ramp2*), up to 32.85 % indel rate in embryo 3. Embryo 2 showed very high off-target editing at sites 3 and 4 (58 % and 99.4 %, respectively), with editing at site 4 as high as on-target (99.4 %), indicating that off-target editing at this site occurred at a very early cellular stage.

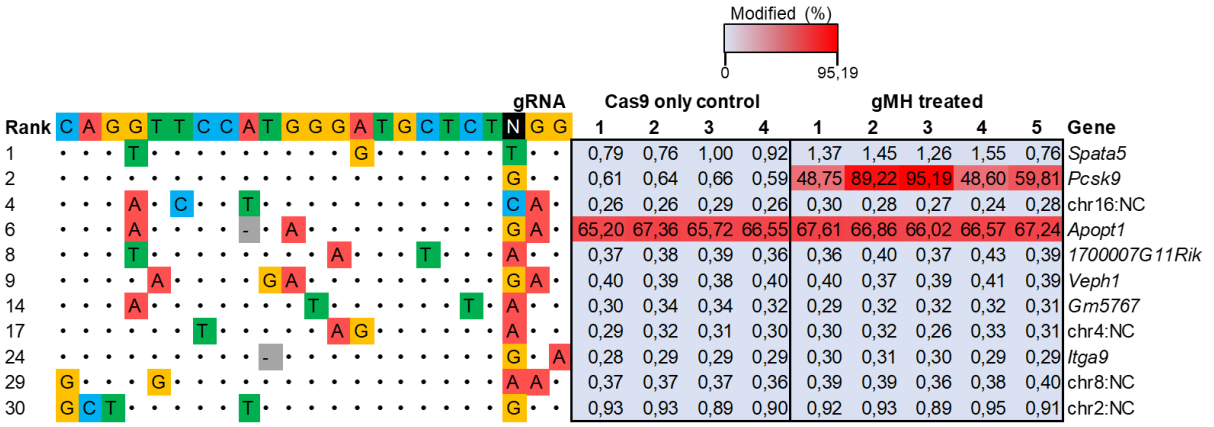

Figure 4.1. Assessment of off-target indels induced by gMH-SpCas9. Indel frequencies measured by targeted amplicon sequencing for the gMH on-target site (rank 2) and 11 off-target sites determined by CIRCLE-seq are shown as heat maps[5]. Each locus was assayed in 4 control embryos injected with Cas9 only and 5 embryos injected with both Cas9 and gRNA. Mismatches to the on-target site are shown as colored boxes. NC: non-coding.

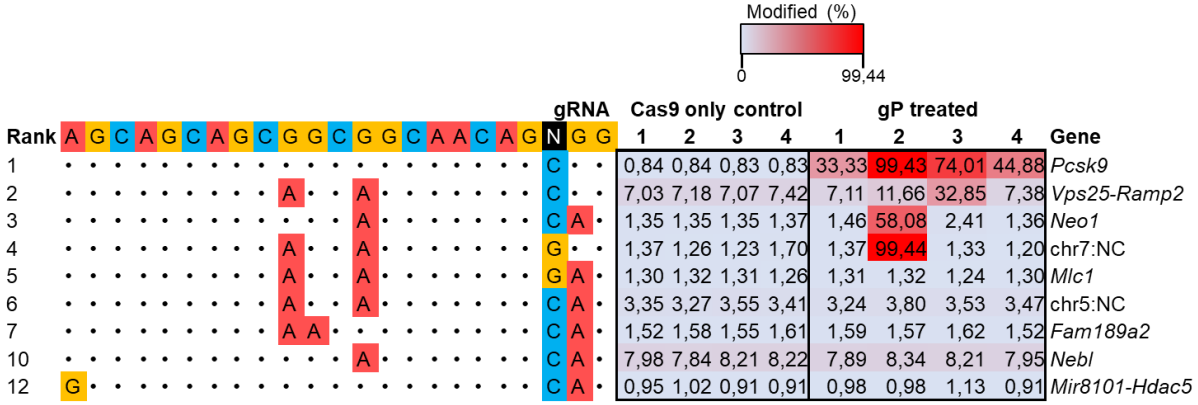

Figure 4.2. Assessment of off-target indels induced by gP-SpCas9. Indel frequencies measured by targeted amplicon sequencing for the gP on-target site (rank 1) and 8 previously validated off-target sites are shown as heat maps[5]. Each locus was assayed in 4 control embryos injected with Cas9 only and 4 embryos injected with both Cas9 and gRNA. Mismatches to the on-target site are shown as colored boxes. NC: non-coding.

Targeted amplicon sequencing would not be able to detect gross chromosomal changes caused by Cas9 editing. In order to evaluate possible translocation events between on-target and off-target sites upon Cas9 induced DSBs, PCR amplification was performed using one primer targeting the on-target region and another primer targeting an off-target region, amplifying the hypothetical translocation event. Amplification products were visualized on Fragment Analyzer and are shown in *Figure 4.3*. Expected translocation product sizes were determined using the genomic location of the primer binding sites and predicted site of Cas9 DSB, and are shown as colored bands corresponding to each translocation event. All primer combinations showed many different unspecific PCR products, but none of the predicted size, with the exception of site 1-site 8 (coded yellow in the lower gel). Bands with the predicted size (253 bp) are visible in 5 out of 8 embryos, but since the bands are present and even more frequent in controls (3 controls versus 2 positive) these products are not likely due to any translocation event induced by Cas9.

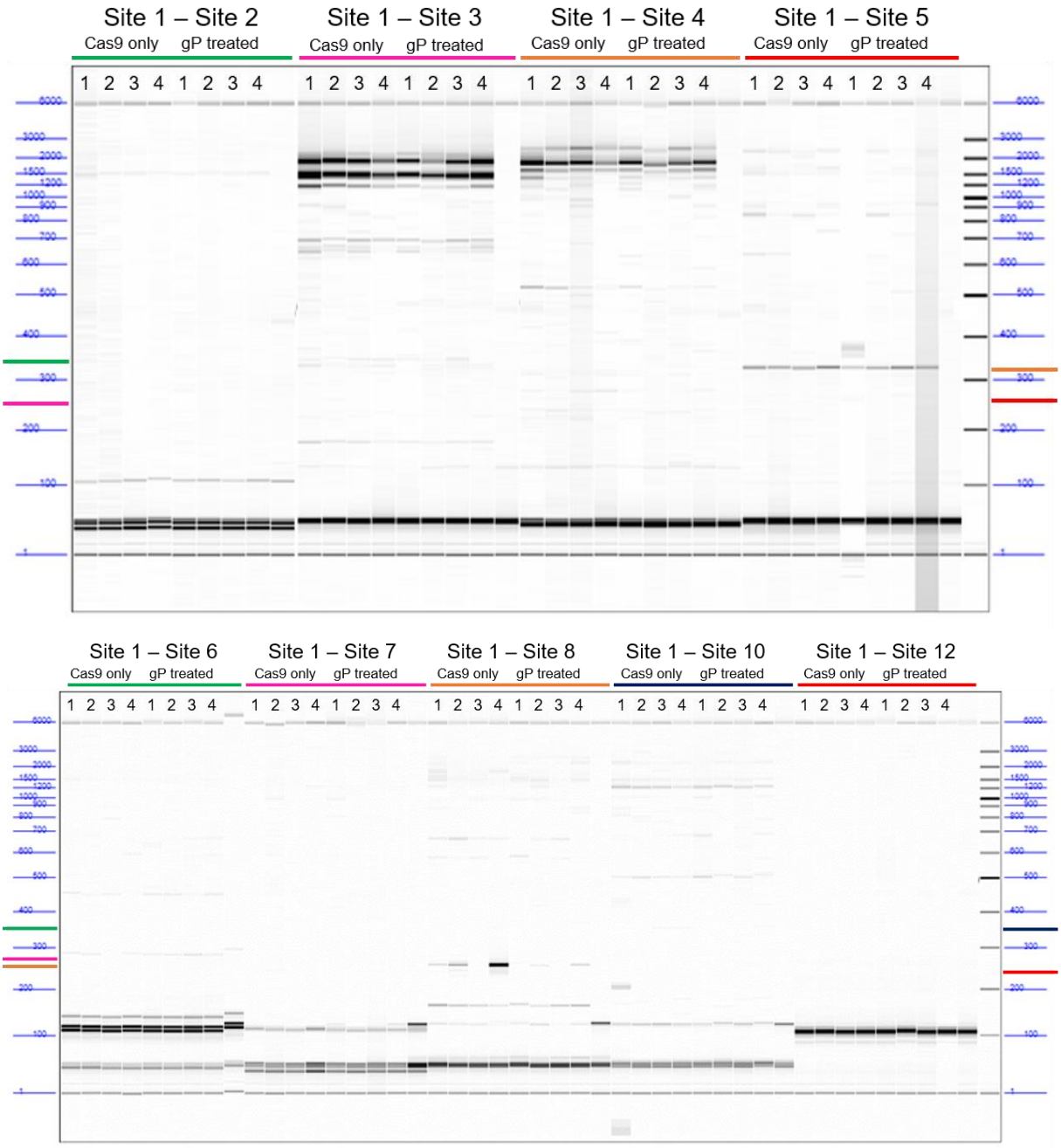

Figure 4.3. Assessment of translocation events between on-target and off-target sites induced by gP-SpCas9. Fragment Analyzer gel image of PCR products using one primer binding to on-target region and another primer binding to off-target region, amplifying the expected translocation event. Translocation events are color-coded, with corresponding expected translocation product size marked as a colored band in-between the ladder bands.

4.2 AzCas9 editing in mouse cells

Evaluation of editing activity of AzCas9 with gRNAs targeting *Pcsk9* and *Nfx1* genes (gP and Nfx) in mouse cells was performed by transfection of Neuro2a cells. A protocol with high transfection efficiency was established, assessed by fluorescence microscopy (example with representative results shown in *Figure 4.4*). Up to 95 % efficiency was achieved with a control GFP plasmid, and between 70-90 % efficiency was generally reproducible with Cas9 variants co-expressed with GFP.

Although comparable transfection efficiencies of SpCas9-GFP and AzCas9 were achieved, total expression of SpCas9-GFP in Neuro2a cells was noticeably lower. Variations in the delivery vectors that might affect expression were found, including different Kozak consensus sequences, number and placement of coding sequences other than Cas9-GFP and overall size of the plasmid. Since the expression of AzCas9-GFP was higher, a standardized delivery vector for SpCas9-GFP, based on the AzCas9-GFP plasmid, was constructed and tested in cells. The new SpCas9-GFP vector showed expression equivalent to AzCas9-GFP, but no difference was observed in the on-target editing activity.

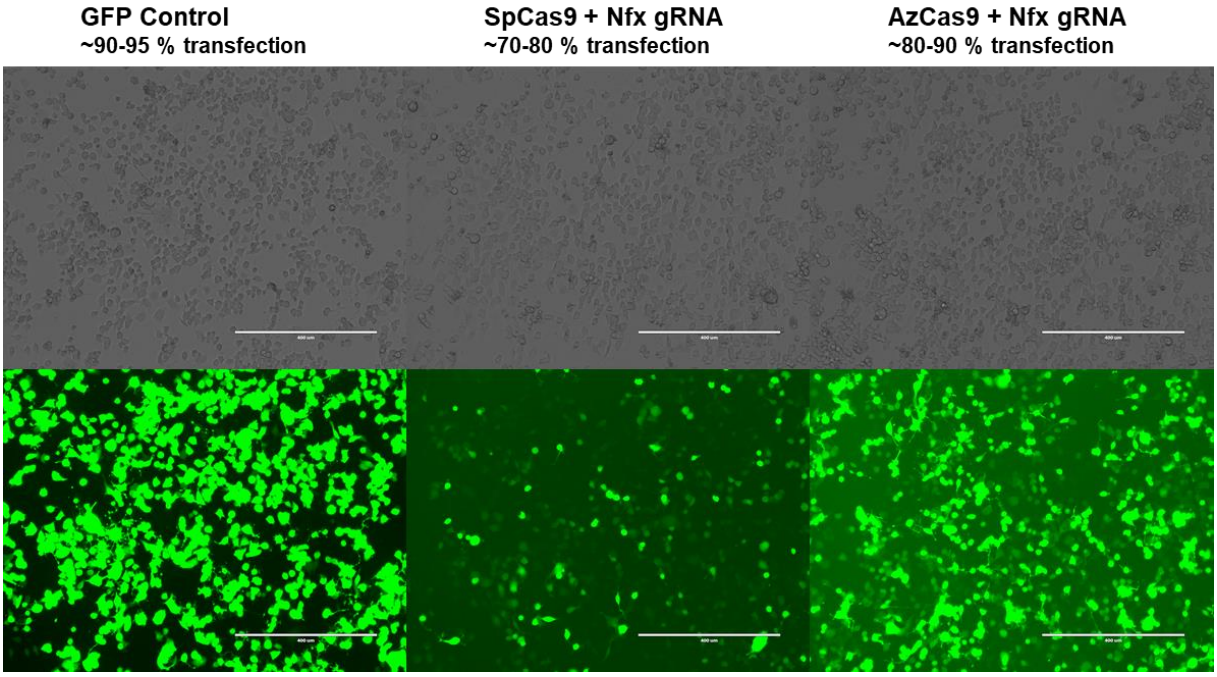

Figure 4.4. Assessment of transfection efficiency of Cas9-GFP variants in mouse Neuro2a cells using fluorescence microscopy.

4.2.1 Editing at the Nfx site

Indel frequencies of Cas9-GFP variants with the Nfx gRNA and initial dsODN donors in Neuro2a cells were measured by targeted amplicon sequencing, and are shown in *Figure 4.5*. A consistently high editing efficiency at around 65 % was achieved with SpCas9-GFP. An average efficiency at around 35 % was achieved for AzCas9-GFP. Addition of the initial dsODN donors did not significantly affect the overall editing efficacy.

After initial transfection experiments, a second batch of dsODN donors were designed. In order to increase stability of oligo duplexes during and after transfection, oligos were designed containing phosphorothioate linkers and 5'-phosphates on the ends (sequences are available in *Appendix B*). Indel frequencies of Cas9 variants with the Nfx gRNA and the second batch of dsODN are shown in *Figure 4.6*. Average editing efficiencies was maintained for both SpCas9 and AzCas9 using dsODN without phosphorothioate linkers (around 65 and 35 %, respectively). Indel frequencies dropped significantly when phosphorothioated dsODN were co-transfected, for both Cas9 variants. Indel frequencies dropped even further when a 5'-phosphate was added to phosphorothioated dsODN.

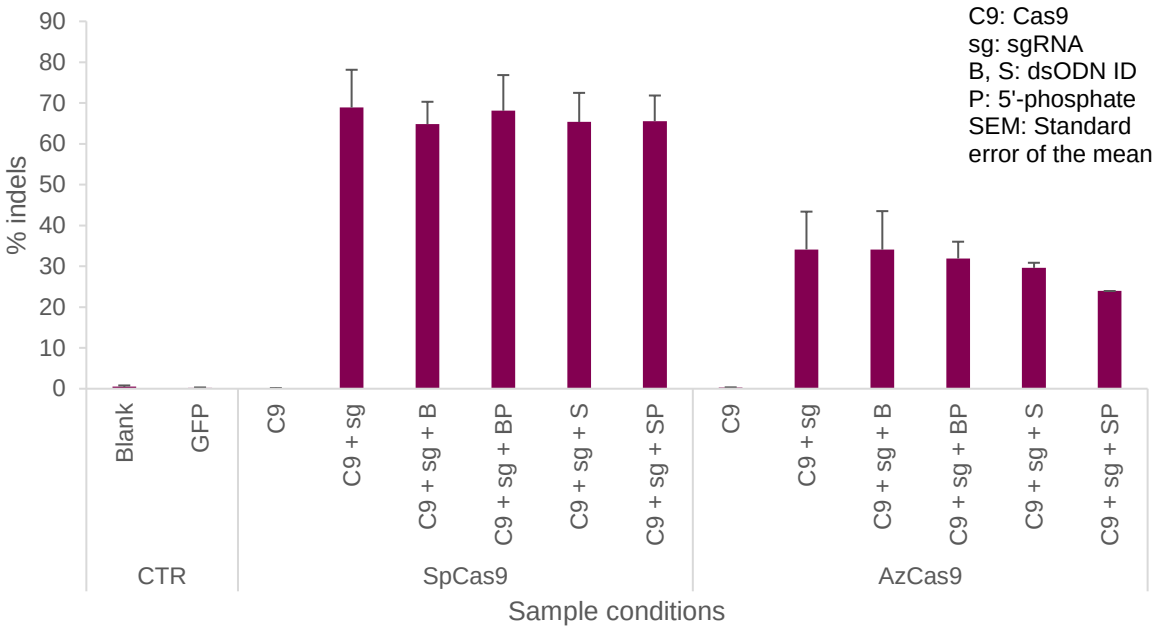

Figure 4.5. Indel frequencies of Cas9 variants with Nfx gRNA and first batch of dsODN in mouse Neuro2a cells. Data is presented as the average of n=2 independent experiments, with error bars representing SEM. Each transfection condition shows the combination of Cas9, gRNA and dsODN insert that was co-transfected.

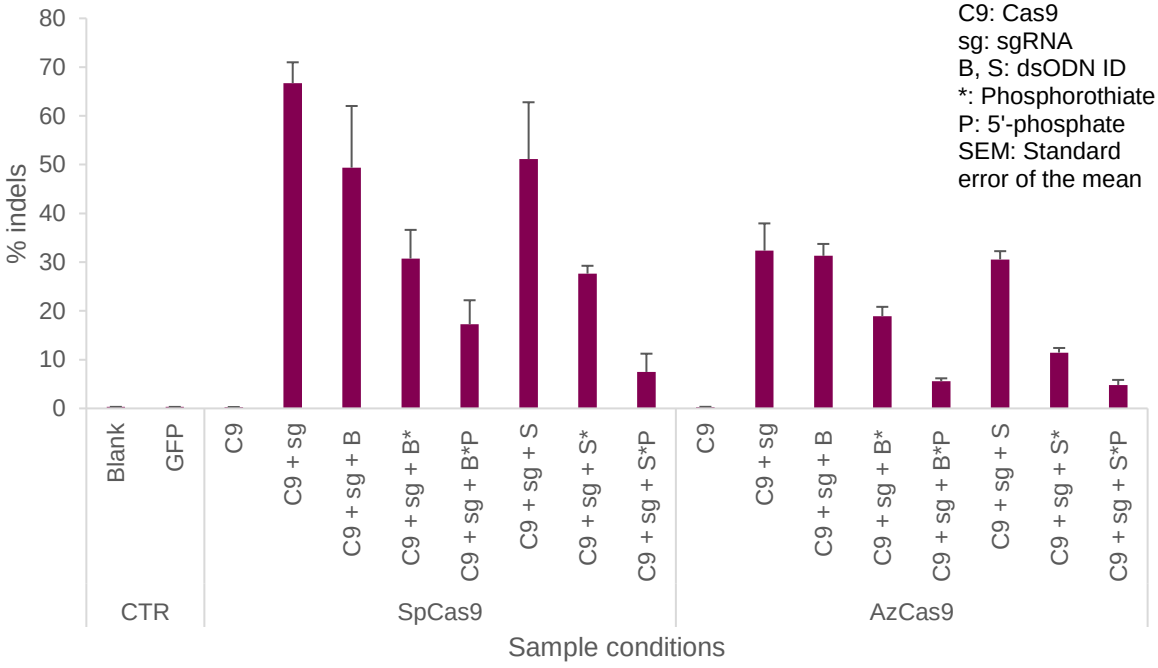

Figure 4.6. Indel frequencies of Cas9 variants with Nfx gRNA and second batch of dsODN in mouse Neuro2a cells. Data is presented as the average of n=3 independent experiments, with error bars representing SEM. Each transfection condition shows the combination of Cas9, gRNA and dsODN insert that was co-transfected.

4.2.2 Editing at the gP site

Initial NGS results of Neuro2a cells treated with Cas9 variants and the gP gRNA showed more than 99 % indels in all samples, including untreated controls (data not shown). Looking closer at the sequences of individual reads, it was discovered that reads mapping to the *Pcsk9* gene contained a 9 bp deletion at the precise location of the gRNA on-target site. The results were confirmed by Sanger sequencing the samples, shown in *Figure 4.7*. Since the relatively large

gap is precisely in the middle of the target site, the gRNA cannot recognize the sequence and induce Cas9 cleavage.

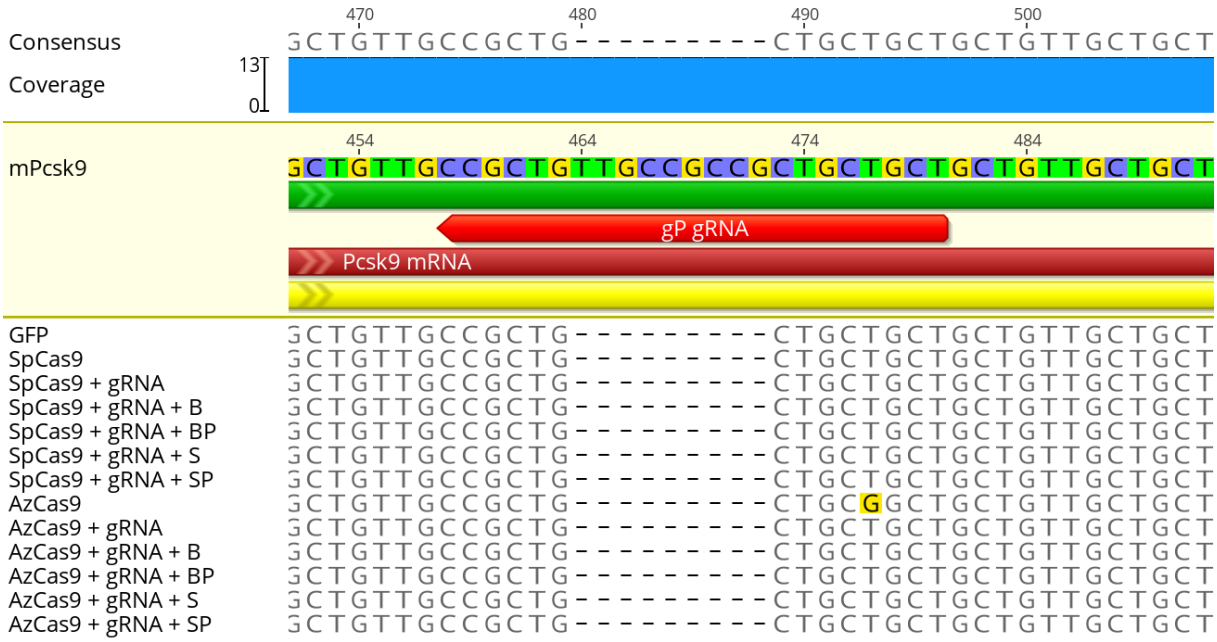

Figure 4.7. Sanger sequencing of the gP on-target site of Cas9 treated Neuro2a cells. The gP gRNA binding site on the *Pcsk9* gene (green) is highlighted in red. Image taken from Geneious version 2019.1 created by Biomatters. Available from https://www.geneious.com.

Initially we thought that the Neuro2a cells used had been somehow contaminated by cells already edited, because of the improbability of such a large deletion at the exactly the site where the gRNA was supposed to bind. Frozen Neuro2a vials from many different sources (including the AstraZeneca UK cell bank) were retrieved, cultured and sequenced. Sequencing revealed that all Neuro2a cells had the same genotype, indicating that the 9 bp deletion was a cell line specific genetic variation. Thus, in order to test the gP gRNA in mouse cells a new cell line was required.

A pilot transfection experiment was run in NIH/3T3 cells, since this cell line had previously been shown to be easily transfected. The cells were transfected with Cas9-GFP variants and gRNA plasmids, using Lipofectamine 3000 Reagent (ThermoFisher) according to the manufacturer's instructions. T7EI assay was run to quickly establish if the gP gRNA showed any editing in these cells (*Figure 4.8*). The Nfx gRNA was used as a positive control, since it had already been proven efficient in Neuro2a cells. The T7EI assay showed substantial editing for both SpCas9 and AzCas9 using the Nfx gRNA, as can be seen from the bands present at the expected cleavage product lengths in samples treated with gRNA. The gP gRNA showed no difference between controls and positive samples, thus indicating no editing in any of the samples. These results were later confirmed by targeted amplicon sequencing, where again all gP site samples showed around 99 % indels and the same 9 bp deletion at the target site. The same genotype was confirmed by Sanger sequencing.

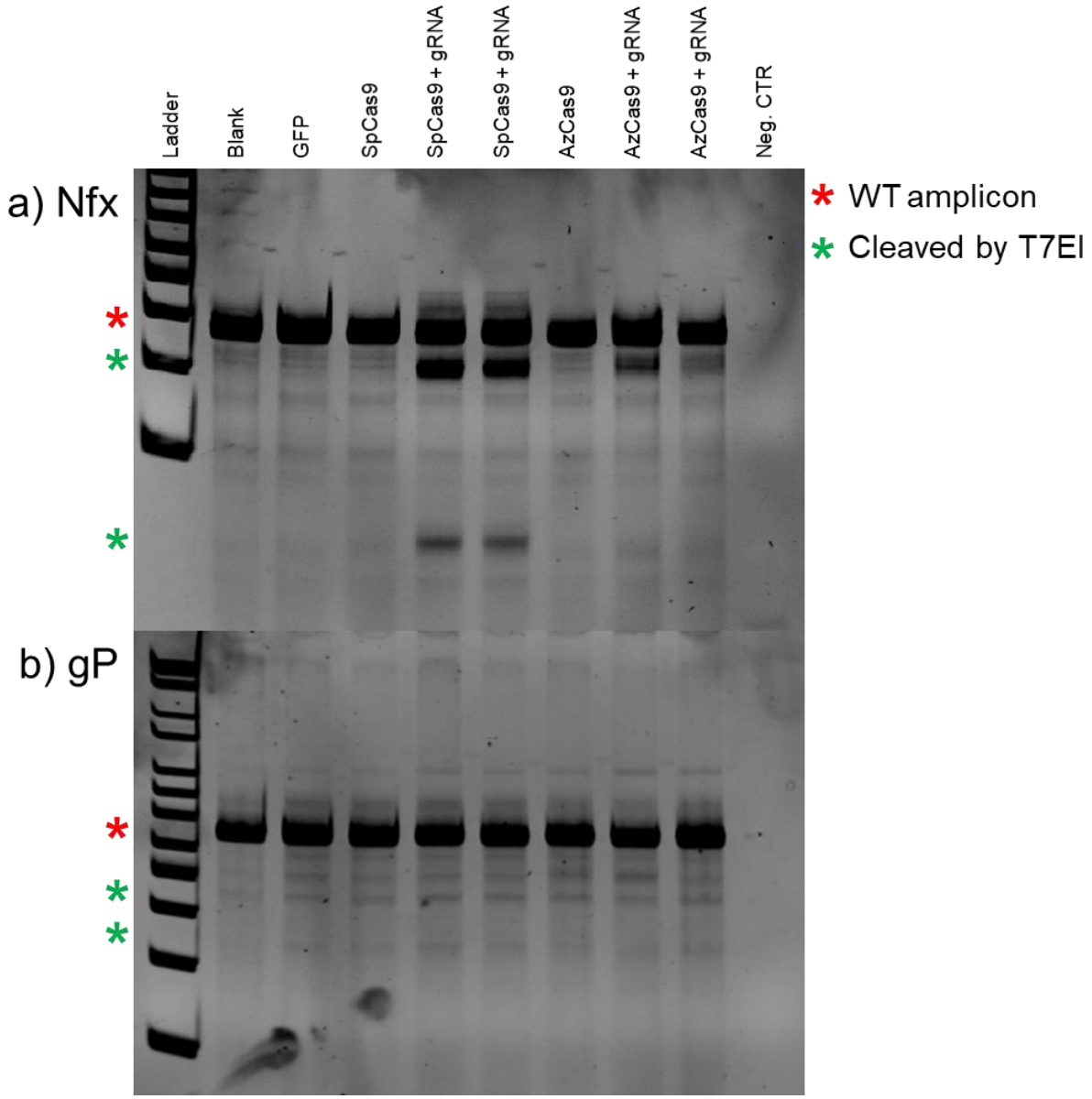

Figure 4.8. T7EI assay on Cas9 treated NIH/3T3 cells. Top gel contains samples treated with the Nfx gRNA and Cas9 variants, bottom gel contains samples treated with gP gRNA and Cas9. Expected product bands are marked by asterisks.

In order to find a cell line with the correct gP on-target site, the gP target site was amplified and sequenced using genomic DNA from many different mouse cell lines and tissues. Liver tissues from C57BL/6N laboratory mice, AML12 and primary mouse embryonic fibroblast (MEF) cells derived from C57BL/6N were all found to have the correct sequence.

Once the correct gP target site sequence was confirmed, a pilot transfection experiment was performed in primary MEF cells, using FuGENE HD. Cas9 and gRNA treated samples were sequenced by targeted amplicon sequencing. Editing was observed in gRNA treated samples, for both SpCas9 and AzCas9 using both gP and Nfx (*Table 4.1*). SpCas9 was found to have increased activity with both guides, as compared to AzCas9. Two different versions of the AzCas9 tracrRNA scaffold were tested for both guides. The second version (v2) had previously been shown to increase the editing activity of AzCas9 in human cells. Editing at the gP target site was slightly higher using the original version (v1) of the scaffold, 8.3 % as compared to 6.6

% with v2. AzCas9 editing rates at the Nfx target site was very similar using both versions of the scaffold, with both at 3.5-3.6 % editing.

Table 4.1. Indels induced by Cas9 variants in primary mouse embryonic fibroblast cells.

Target site	Sample	% Reads with indel
gP	Blank	0,0
	GFP	0,0
	SpCas9	0,0
	SpCas9 + gRNA	14,3
	AzCas9	0,0
	AzCas9 + gRNA v1	8,3
	AzCas9 + gRNA v2	6,6
Nfx	Blank	0,0
	GFP	0,0
	SpCas9	0,0
	SpCas9 + gRNA	18,4
	AzCas9	0,0
	AzCas9 + gRNA v1	3,5
	AzCas9 + gRNA v2	3,6

Transfection efficiency in MEF cells using FuGENE HD was relatively low, and appeared to be cytotoxic. Fluorescence microscopy after transfection showed that many successfully transfected cells expressing Cas9-GFP exhibited a changed morphology, either round and detached from the surface or with the extracellular membrane burst open (example with representative results shown in *Figure 4.9*). A protocol for electroporation of MEF cells using the Neon Transfection System (ThermoFisher, Cat. No. MPK5000), previously used for delivery of ribonucleoproteins (RNP) with high efficiency, was tried as an alternative. However, transfection of Cas9 expression plasmids using this protocol did not yield the same efficacy, possibly due to the large size of the plasmid vectors.

4.3 CIRCLE-seq

CIRCLE-seq experiments for AzCas9 with gP are currently ongoing. Results are forthcoming.

4.4 AzCas9 ObLiGaRe donor vector

No successful assembly of the AzCas9 ObLiGaRe donor vector has yet been achieved. AzCas9 ObLiGaRe vector generation is currently ongoing.

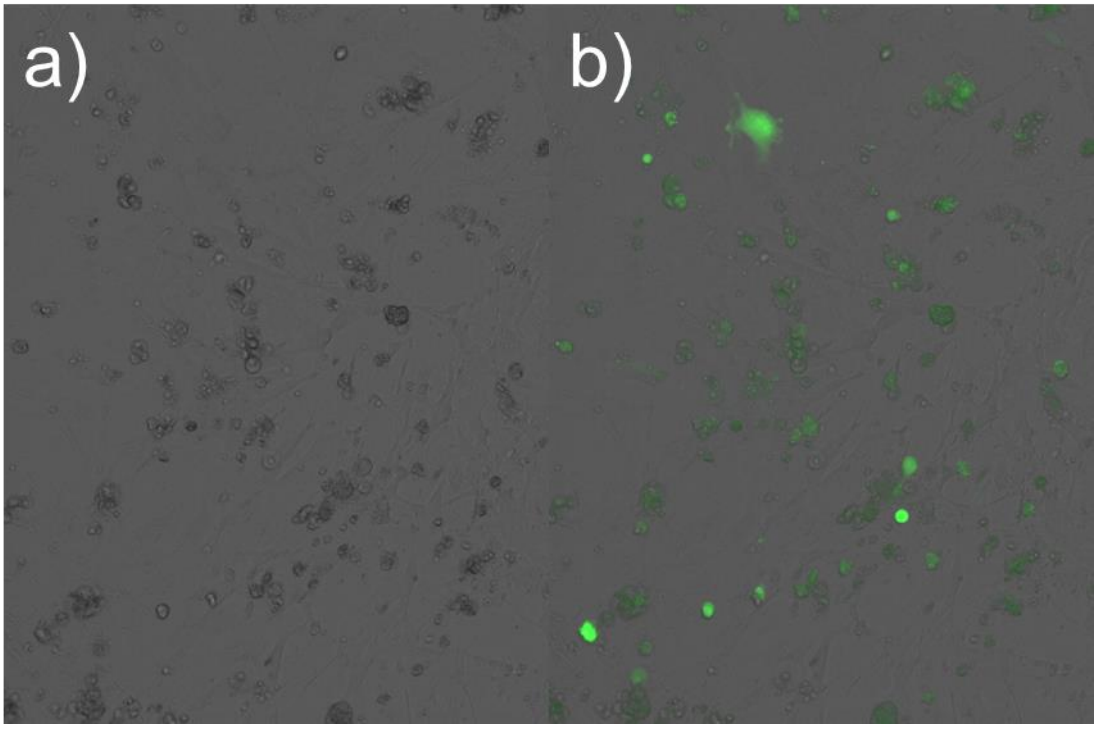

Figure 4.9. Assessment of FuGENE HD transfection efficiency in primary MEF cells using fluorescence microscopy. Cells were transfected with AzCas9-GFP and gP gRNA v2. a) Brightfield microscopy only. b) Brightfield microscopy, overlaid with GFP fluorescence.

5 Discussion

5.1 Off-target editing in embryos

Akcakaya & Bobbin et al. previously described two gRNAs used for *Pcsk9* gene editing in mouse livers delivered via intravenous tail-vein injection: One with no detectable off-targets (gMH), another with many off-targets (gP), with on-target editing efficacies ranging from 21-51 and 23-30 %, respectively[5].

One limitation with current off-target detection methodologies is the detection limit of next generation sequencing, which is 0.1%. This means if there is an off-target editing with frequency less than in one cell among thousand, we would not be able to detect it. Theoretically, gMH may cause off-target mutations that are below the detection limit of next generation sequencing in liver. Gene editing in a single cell would enhance on- and off-target editing efficacy up to 100% depending on the cellular stage that Cas9 acts. Here, we have explored if enhancing the on-target editing efficacy in mouse embryos would reveal any off-targets of gMH previously not identified in livers.

On-target editing with gMH in all five treated embryos was comparable or higher than in livers, ranging from 48.6-95.2 %. Embryos with nearly 100% on-target editing is possibly edited only once, during the single-cell stage, leading to an indel which is inherited by subsequent generations of cells, whereas embryos with lower editing rates might be edited at later stages. Despite achieving up to 95.2 % editing, no mutations were detected at off-target sites, demonstrating the importance of proper gRNA design in improving the specificity of Cas9 editing.

In the previous work by Akcakaya & Bobbin et al., on- and off-target editing were analyzed in the genomic DNA extracted from ten mg liver that is containing 2-3 million cells[32]. It is not known if the edited single cells in the population exhibit differences in their off-targets. To our

knowledge, there is no study reporting individual off-target profiles in single cells, using a gRNA deliberately designed to cause many off-targets. Here, we have explored whether gP induces identical off-targets in all single cells (in this case embryos), and whether the off-target mutations are induced at different time frames during cell division.

On-target editing with gP in all four treated embryos was comparable or higher than in livers, ranging from 33.3-93.4 %. Notably, the promiscuous gP gRNA previously exhibited off-target editing activity in all of the evaluated sites in the livers of all gP treated mice replicates. Strikingly, Cas9 editing profile of individual embryos was found to be somehow distinct. Only two out of four gP treated embryos showed any off-target activity, and only in the three highest ranked off-target sites. The two embryos showing detectable off-target editing correspond to the highest on-target editing rates, indicating that high on-target efficiency is correlated with off-target activity. Notably, none of the embryos exhibited off-target editing in more than three sites, which may indicate that embryos with high number of off-target mutagenesis were incompatible with life.

Cas9 editing activity is likely to be affected by the developmental stage of the embryo at the point of injection. Which chromosomal DNA is available and which cells take up the Cas9 and guide mRNA are factors that could play a role. One possible reason for this is the limitations that come with the method of delivery, namely manually microinjecting the embryos with Cas9 mRNA and gRNA. Manual microinjection is laborious, time-consuming and generally results in low rates of positive samples. Although great care was taken to ensure that embryos were injected at the same time point, the sequencing still showed large differences in on-target editing between embryos, indicating that editing occurred at different developmental stages. Differences may be the result of Cas9 mRNA requiring protein translation to initiate cleavage, and therefore acts at various time points depending on the cellular stage and may cause diverse editing rates.

We have repeated the experiment with a recently developed in-house method of delivering gene editing technologies to embryos via electroporation. Electroporation can be used to deliver Cas9 protein and gRNA in the form of pre-complexed ribonucleoproteins (RNP), to large numbers of embryos simultaneously. By delivering Cas9 and gRNA as an RNP, editing can rapidly occur in target cells, without depending on the translation of Cas9 protein through mRNA. This allows instant and more synchronized editing across the embryos. Off-target analysis in electroporated embryos is ongoing.

5.2 AzCas9 editing in mouse cells

Prior to evaluating the specificity of AzCas9 in the mouse genome using CIRCLE-seq, it is essential to identify gRNAs that work efficiently with the Cas9 variant being examined. For this purpose, we have tested the editing of AzCas9 with two gRNAs, Nfx and gP, in mouse cells. In some experiments, cells were co-transfected with dsODN donors to evaluate knock-in after gene editing occurred. We have evaluated knock-in with duplexed oligonucleotide donors with varied structures. We have also evaluated oligos with modifications that might help improve knock-in, including oligos containing 5'-phosphates and phosphorothioate linkers. Endogenous cellular DNA ligases use the 5'-phosphate ends of the donor to ligate the fragment into the DSB, and phosphorothioate linkers help protect the DNA duplexes from degradation by endogenous cellular exonucleases.

5.2.1 Editing at the Nfx site

Editing efficiency was successfully assessed at the Nfx locus in mouse Neuro2a cells. Cells were co-transfected with Cas9, gRNA and dsODN inserts with varied structures, and editing outcomes were evaluated by targeted deep sequencing. A consistent, high transfection

efficiency was achieved in all conditions. An average editing rate of around 65 % for SpCas9 and around 35 % for AzCas9 was achieved when no dsODN was co-transfected.

Co-transfecting cells with initial dsODN inserts did not significantly affect the indel rate. Indel rates for both Cas9 variants were significantly decreased when adding dsODN with phosphorothioate linkers. Although the linkers might theoretically increase the knock-in rates, exposing cells to these undegradable DNA duplexes can be toxic. This toxicity might enrich the cell populations for untransfected cells, leading to seemingly lower indel rates when evaluating the editing outcomes. Adding 5'-phosphates to phosphorothioated dsODN significantly lowered indel rates even further.

Since the start of transfection experiments, work on optimizing the gRNA scaffold for AzCas9 has been on-going. AzCas9 is a completely novel Cas9 variant, and the design of its tracrRNA was started from scratch by members of the PGE group. Optimization experiments have found a new version of the scaffold (labelled as version 2 in primary MEF cell experiments) showing higher on-target activity in human cells. Targeted deep sequencing of MEF cells treated with either version 1 or version 2 of the Nfx gRNA scaffold showed only a slight improvement in editing, with 3.5 % and 3.6 % editing for each scaffold, respectively.

5.2.2 Editing at the gP site

It was essential to verify that AzCas9 works with gP in a cellular context before CIRCLE-seq experiments. gP is currently the only published promiscuous gRNA targeting the mouse genome, and its off-target profile with SpCas9 has already been established by CIRCLE-seq and verified *in vivo*[5].

An unexpected sequence variation in the gP on-target site was found for multiple mouse cell lines. A 9 bp deletion was present in both Neuro2a and NIH/3T3 cells, located exactly at the gRNA binding site, making editing by both SpCas9 and AzCas9 with gP in these cells impossible. Primary MEF cells were shown to have the correct on-target sequence and were successfully edited at the gP site with both SpCas9 and AzCas9, allowing us to begin the CIRCLE-seq experiments. Two versions of the AzCas9 gRNA scaffold were again tested, with version 1 shown to be significantly more efficient, with 8.3 % editing as compared to 6.6 % with version 2.

Although sufficient editing was achieved to verify that gP works with AzCas9, a consistently high transfection efficiency should ideally be achieved before drawing conclusions about the gRNA efficacies and on/off-target specificities in mouse cells. Alternative cell lines with correct on-target sequences for both gP and Nfx, for efficient plasmid transfection are currently being investigated, including the B16-F10 mouse melanoma cell line which has previously been successfully edited with gP[16]. Once a suitable alternative has been established, a more thorough evaluation of AzCas9 editing using both scaffold versions of gP and Nfx will be performed.

5.3 CIRCLE-seq

CIRLCE-seq experiments are currently on-going, a discussion of the results is forthcoming.

5.4 AzCas9 ObLiGaRe donor vector

No successful assembly of the AzCas9 ObLiGaRe donor vector has yet been achieved. The correct fragments for cloning have been produced and purified, but transformation after the Gibson assembly reaction has yielded no bacterial colonies thus far. The reaction and subsequent transformation have been tried several times using various backbone:insert ratios, and several controls have been implemented to troubleshoot the problem.

The reaction was initially tried using 30 bp homology between fragments. After digestion with SwaI, the resulting linear fragment contains 4 bp left over from the RE site on either end. These 4 bp were not included in the original primers, causing 4 bp of non-homology between the backbone and the insert fragment. Previous reactions generating ObLiGaRe donor vectors had found this non-homology not to affect the efficiency of the reaction. The transformations using 30 bp homology between fragments yielded no colonies.

Assembly reactions were run on agarose gel to identify the problem. Efficient Gibson assemblies, such as the positive control included in the NEB Gibson Assembly Master Mix, should show individual fragments disappearing after the reaction. Gels of donor vector assembly reactions showed most of the fragments still being present, indicating low reaction efficiency. Gels also showed a longer, linear fragment which could be the final product in a linear form.

To increase efficiency, two new primer pairs were designed. These included 50 bp homology to the linear backbone, and either with or without the 4 bp homology to the left-over SwaI site. New fragments were generated by PCR and used in new assembly reactions together with the digested backbone. New assembly reactions did not yield any colonies so far. Agarose gels of the reactions show low overall reaction efficiencies.

References

1 Doudna, J. A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science (New York, N.Y.)* **346**, 1258096, doi:10.1126/science.1258096 (2014).

2 Koonin, E. V., Makarova, K. S. & Zhang, F. Diversity, classification and evolution of CRISPR-Cas systems. *Current opinion in microbiology* **37**, 67-78, doi:10.1016/j.mib.2017.05.008 (2017).

3 Cebrian-Serrano, A. & Davies, B. CRISPR-Cas orthologues and variants: optimizing the repertoire, specificity and delivery of genome engineering tools. *Mammalian genome : official journal of the International Mammalian Genome Society* **28**, 247-261, doi:10.1007/s00335-017-9697-4 (2017).

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

5 Akcakaya, P. *et al.* In vivo CRISPR editing with no detectable genome-wide off-target mutations. *Nature* **561**, 416-419, doi:10.1038/s41586-018-0500-9 (2018).

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

7 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, doi:10.1101/gr.145441.112 (2013).

8 Sadelain, M., Papapetrou, E. P. & Bushman, F. D. Safe harbours for the integration of new DNA in the human genome. *Nature Reviews Cancer* **12**, 51, doi:10.1038/nrc3179 (2011).

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

10 Heigwer, F., Kerr, G. & Boutros, M. E-CRISP: fast CRISPR target site identification. *Nature methods* **11**, 122-123, doi:10.1038/nmeth.2812 (2014).

11 Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nature Biotechnology* **33**, 187, doi:10.1038/nbt.3117 (2014).

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

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

14 Kim, D. *et al.* Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nature methods* **12**, 237, doi:10.1038/nmeth.3284 (2015).

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

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

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

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

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

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

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

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

23 Hirano, H. *et al.* Structure and Engineering of Francisella novicida Cas9. *Cell* **164**, 950-961, doi:10.1016/j.cell.2016.01.039 (2016).

24 Abudayyeh, O. O. *et al.* RNA targeting with CRISPR–Cas13. *Nature* **550**, 280, doi:10.1038/nature24049 (2017).

25 Kleinstiver, B. P. *et al.* Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. *Nat Biotechnol* **33**, 1293-1298, doi:10.1038/nbt.3404 (2015).

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

27 Yin, H., Kauffman, K. J. & Anderson, D. G. Delivery technologies for genome editing. *Nature reviews. Drug discovery* **16**, 387-399, doi:10.1038/nrd.2016.280 (2017).

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

29 Sentmanat, M. F., Peters, S. T., Florian, C. P., Connelly, J. P. & Pruett-Miller, S. M. A Survey of Validation Strategies for CRISPR-Cas9 Editing. *Scientific Reports* **8**, 888, doi:10.1038/s41598-018-19441-8 (2018).

30 Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods* **6**, 343, doi:10.1038/nmeth.1318 (2009).

31 Pinello, L. *et al.* Analyzing CRISPR genome-editing experiments with CRISPResso. *Nature Biotechnology* **34**, 695, doi:10.1038/nbt.3583 (2016).

32 Klaunig, J. E. *et al.* Mouse liver cell culture. I. Hepatocyte isolation. *In vitro* **17**, 913-925 (1981).

Appendix A – Detailed laboratory protocols

This appendix provides detailed, step-by-step laboratory instructions for all experiments, as well as motivations and clarifications for the selected methods.

Targeted deep sequencing

This general workflow was used for all targeted deep sequencing, including editing evaluation after transfection experiments. Below is the workflow described for sequencing on/off-target sites of Cas9 treated embryos.

1st PCR.

An initial PCR program was run to amplify the region of interest i.e. the Cas9 on/off-target cleavage site. Genomic DNA extracted from 10 mgs mouse embryos injected with Cas9 mRNA and either of the two gRNAs, gP and gMH, were provided by P. Akçakaya.

All sites were amplified using Q5 High Fidelity polymerase (NEB, Cat. No. M0492L), from 75-150 ng of input genomic DNA using site-specific primers attached with sequencing adapters:

5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-[amplicon specific primer]-3' and 5'-GTCTCGTGGGCTCG GAGATGTGTATAAGAGACAG-[amplicon specific primer]-3`

The sequencing adapters are used in a subsequent PCR reaction, to anneal a combination of indexing primers that will produce barcoded samples. A list of all primers used is provided in *Appendix B*.

A typical master mix and PCR cycling numbers are shown below. Products were verified on 1% agarose gels. If the gel indicated primer dimer formation, the PCR was repeated with halved primer concentrations. If primer dimers were still present, primer concentration was further reduced.

In total, 9 sites from embryos injected with gP and 11 sites from embryos injected with gMH were successfully amplified. DNA from embryos injected with only Cas9 was used as negative control for all sites. DNA from 4 embryos treated with only Cas9, 4 embryos treated with Cas9 and gP, and 5 embryos treated with Cas9 and gMH was amplified for each on- or off-target site.

1st PCR master mixes and cycling numbers.

Q5 Master Mix			**PCR cycles with Q5 High Fidelity**		
Reagent	**Vol. 1x**		**Temp**	**Time**	**Cycles**
Q5 High Fidelity 2x	7.5		98° C	30 sec	
NGS Primer up (10 µM)	0.75		98° C	10 sec	Optional: X5
NGS Primer low (10 µM)	0.75		82° C	20 sec	
Nuclease free water	5		72° C	10 sec	
Total µL MM	14		98° C	10 sec	X30
Aliquot in PCR strips, add DNA			T_m° C	20 sec	
			72° C	20 sec	
DNA (100 ng/µL)	1		72° C	2 min	
Total rxn vol	15		4° C	For ever	

PCR clean-up.

PCR products were purified and size-selected using Agencourt AmPure XP magnetic beads (Beckman Coulter, Cat. No. A63880). AMPure XP binds fragments 100 bp and above (the amplicons) and excludes primers 50 bases and below. Procedure:

1. Add nuclease free water to PCR sample in PCR strip to a final volume of 50 µL.
2. Vortex room temperature beads for 20 seconds and pour onto a reservoir.
3. Add 90 µL beads to each sample (1:1.8 ratio).
4. Pipette mix gently and well at least 10x (a good mixing helps the DNA to bind to the beads).
5. Incubate 10 min at room temperature.
6. Place strip on a magnetic plate and incubate 5 min or until clear.
7. Remove and discard the supernatant (when clear) but leave the last 5 µL to not disturb the beads.
8. While the strip is on the magnetic plate, add 200 µL 80% EtOH without disturbing the beads. Incubate for 30 seconds and then remove *all* EtOH while the beads are on the magnet.
9. Do the wash for total of two times.
10. Remove your strip from the magnet.
11. Air dry the beads until all the ethanol is gone and the beads are not shiny anymore. Be careful to not let them dry too much and crack, since it will lead to a lower DNA yield.
12. Add 26 µL elution buffer (Qiagen Buffer EB, Cat. No.19086) and pipette mix 10x. Incubate 5 min at room temperature.
13. Place the strip back on the magnet and incubate for 2 min or until clear.
14. Transfer 25µL (In two transfers, 12.5+12.5 µl) eluate to a new tube without transferring any beads.
15. After purification, run 2 µL of your sample on Fragment Analyzer (DNF-915 kit, Advanced Analytical Technologies) to verify product. Ensure correct size, specific product and no primer-dimers.

Dilution and pooling.

Multiple sites originating from the same embryo were diluted and pooled for the second PCR. Multiple sites can be pooled at this stage because they originate from different genomic locations, the NGS reads are easily sorted after mapping as long as the location of the site is known. Purified samples were diluted to 0.67 ng/µl in 35 µl Buffer EB (Qiagen, Cat No. 19086), using the estimated concentrations of samples from Fragment Analyzer. 3-5 sites from the same embryo were pooled together for the 2nd PCR.

Example of pooling procedure below:

Pool	Well	Sample ID	ng/uL	10x vol (µL)	Qia EB
1	B1	gP 5-1 site1	51,70	0,45	
	A1	gP 5-1 site2	37,87	0,62	
	C1	gP 5-1 site3	21,98	1,07	
	B1	gP 5-1 site4	13,26	1,77	
	D1	gP 5-1 site5	47,15	0,50	
					170,59

Pool	Well	Sample ID	ng/uL	10x vol (µL)	Qia EB
17	A1	gMH 5-1 site1	13,57	1,73	
	B1	gMH 5-1 site2	38,96	0,60	
	C1	gMH 5-1 site4	33,84	0,69	
					101,98

Indexing PCR.

A second PCR was run on the pools to index the samples. A set of sequencing primers (Illumina Nextera XT index kit, Cat. No. FC-131-1002), annealing to the adapters added in the first PCR, was used to barcode each pool to enable further pooling. Each sample has one N7xx primer and one S5xx primer that is unique for that sample, so that each read can be sorted after sequencing.

All pools were amplified with Kapa HiFi Polymerase (Roche, Cat No. KK2602), using the below PCR master mix and cycling numbers.

Typical PCR master mixes and cycling numbers.

PCR MasterMix KAPA HiFi:	
Reagent	**Vol. 1x (µl)**
KAPA HiFi* 2x	12,5
NGS Primer 1 (N7xx)	2,5
NGS Primer 2 (S5xx)	2,5
Water	6
10xDNA	1.5
Total reaction volume	**25**
*KAPA HiFi Hotstart Ready Mix	

PCR cycles with KAPA HiFi HotStart		
Temp	**Time**	**Cycles**
72 °C	3 min	
98 °C	30 sec	
98 °C	10 sec	X10
63 °C	30 sec	
72 °C	3 min	
72 °C	5 min	
4 °C	For ever	

PCR products were cleaned up using Agencourt AmPure XP beads, same as the previous purification but with a ratio of 1:1 (=50µL beads) this time. Purified products were verified on Fragment Analyzer (DNF-915 kit) to ensure size, concentration and purity.

Normalization and pooling

Fragment Analyzer was used to determine concentration of each pool. Average fragment size and concentration of pools were used to pool and dilute samples to a final molarity of 20 nM

(example seen below), in a single library to be submitted for NGS. Samples were diluted in buffer EB with 0.1% tween 20. The tween helps the very low amount of sample to not attach to the tube walls.

Sample Name	Finished Library Fragment Size (bp)	Finished Library Concentration (ng/µl)	**Amt of Library (µl)**	Amt of EB + Tween (µl)	**Total EB + Tween (µl)**	**Desired Molarity (nM)**	**Total Volume per sample (µl)**
Pool 1	350,00	9,55	**7,26**	7,74	**37,61**	20	15
Pool 2	351,00	9,30	**7,47**	7,53		20	15
Pool 3	351,00	8,95	**7,76**	7,24		20	15
Pool 4	342,00	10,10	**6,71**	8,29		20	15
Pool 5	350,00	8,46	**8,19**	6,81		20	15

1 µL of the finished library was run (in triplicates) on Qubit (dsDNA HS Assay Kit, Cat. No. Q32854) to determine the concentration. This step is very important since the exact concentration of the pool is needed for further dilutions before sequencing.

Final libraries were submitted for sequencing on an Illumina NextSeq500 by the AstraZeneca NGS/Transcriptomics team.

Embryo translocation PCRs

A PCR was run to check for translocation events between the on-target and off-target sites of the gP gRNA. The forward primer for the on-target site (site 1, PA307) was used in conjunction with the reverse primer (even-numbered PA primers) for an off-target site to check for PCR products. A list of all primers used is provided in *Appendix B*. Standard Phusion Flash (ThermoFisher, Cat. No. F-548S) PCR was run with T_a=64°C for all conditions.

PCR Master Mix	
Reagent	**1x Rxn (µl)**
2x Phusion Flash	10
Primer fwd (10µM)	1
Primer rev (10µM)	1
dH_2O	7
DNA (50 ng/µl)	1
Total	20

PCR Cycling Numbers		
Temp (°C)	t (sec)	Cycles
98	10	
98	1	30x
64	5	
72	15	
72	60	
4	For ever	

Cloning gRNA via annealed oligos

1. Digest 1ug of backbone with BsaI-HF (New England Biolabs, Cat. No. R3535) for 3h at 37°C:

Reagent	**Volume**
Backbone DNA	1 µg
10x CutSmart Buffer (NEB, Cat. No. B7204S)	5 µl
BsaI-HF (NEB, Cat. No. R3535)	1 µl (10 units)
Nuclease-free water	x µl (up to 50 µl)
Total	50 µl

2. Run all of the reaction on 1% agarose gel and **purify digested backbone** using QIAquick Gel Extraction Kit (Qiagen, Cat. No. 28706) according to manufacturer's instructions.

3. Anneal each pair of gRNA oligos*:

Reagent	**Volume**
Oligo 1 (100 µM)	1 µl
Oligo 2 (100 µM)	1 µl
Duplex Buffer (IDT)	8 µl
Total	10 µl

*A list of all gRNA oligos used is provided in *Appendix B*.

Anneal 5 minutes at 95 °C, cool down to room temperature.

4. Dilute annealed oligos 100x in nuclease-free water.

5. Set up ligation reaction and incubate at room temperature for 60-90 mins:

Reagent	**Volume**
BsaI digested vector from **step 2**	10 ng
Annealed oligo duplex from **step 4** (100x diluted)	1 µl
10x T4 Ligase Buffer (NEB, Cat. No. M0202S)	1 µl
T4 DNA Ligase (NEB, Cat. No. M0202S)	1 µl
Nuclease-free water	x µl (up to 10 µl)
Total	10 µl

6. Transform into DH5-alpha or DH10-beta (New England Biolabs, Cat. No. C2987H and C3019H) **and plate on kanamycin selection plates:**

1. Thaw cells on ice for 10 mins, gently pipette 25 µl cells to a tube on ice.
2. Add 5 µl of ligation from **step 5** to 25 µl of competent cells, carefully flick the tube 4-5 times to mix.
3. Incubate on ice 30 min. Do not mix.
4. Heat shock 30 sec at 42 °C. Do not mix.
5. Incubate 5 mins on ice. Do not mix.
6. Add 400 µl of room-temperature SOC medium.
7. Recover at 37 °C with 500 rpm shaking in table thermoshaker for 1 h.
8. Plate 100 µl on one plate, 300 µl on another.
9. Grow at 37 °C overnight.

7. Scale-up and verification:

Pick successfully cloned colonies with a pipette tip and add to 4 ml LB medium with 0.1 % kanamycin in falcon tubes. Grow at 37 °C overnight with 500 rpm shaking.

Add 500 µl overnight culture to a microcentrifuge tube and put at room temperature. Purify sgRNA plasmids from the remaining overnight culture using QIAprep Spin Miniprep Kit (Qiagen, Cat. No. 27106) according to the manufacturer's instructions. Verify correct insertion of sgRNA oligos by Sanger sequencing plasmids with BigDye Terminator (see next section). Use 500 µl overnight culture with verified plasmids to inoculate 200 ml LB medium with 0.1 % kanamycin in a 1 l flask. Grow at 37 °C overnight with 500 rpm shaking.

Purify sgRNA plasmids from 200 ml liquid culture with EndoFree Plasmid Maxi Kit (Qiagen, Cat. No. 12362) according to the manufacturer's instructions. Endo-toxin free plasmids are ready for transfection.

Sanger sequencing plasmids with BigDye Terminator

Plasmids were sequenced using the BigDye Terminator v3.1 Cycle Sequencing Kit (ThermoFisher, Cat. No. 4337456). BigDye Terminator creates PCR products spanning 500-700 bps from the primer binding site. PCR products were then washed and sequenced on a 3730 DNA Analyzer instrument (ThermoFisher, Cat. No. 3730S). A list of sequencing primers used for plasmids is not provided, since plasmid sequences were considered proprietary.

Run a BigDye Terminator PCR with a single primer to create sequencing reads:

Reagent	**Volume**
BigDye Buffer	4 µl
BigDye Terminator	0.5 µl
Nuclease free water	14 µl
Primer (10 µM)	0.5 µl
DNA (100 ng/µl)	1 µl
Total	20 µl

PCR Cycling Numbers		
Temp (°C)	t	Cycles
95	2 min	
95	5 sec	35x
55	10 sec	
60	3 min	
4	For ever	

DNA precipitation and wash

Add 5 µl EDTA (125 mM), mix by pipetting up and down 10 times.
Add 60 µl EtOH 100%.
Vortex 20 sec.
Incubate at 4 °C for 20 mins.

Centrifuge at 14.000 rpm for 5 mins (centrifuge for PCR strips).

Place the strip in a 96-well PCR microplate and centrifuge the plate upside-down at 500 rpm for 30 sec with paper tissue covering the open top to dry the EtOH.

Add 60 µl EtOH 70%.
Vortex 20 sec.

Centrifuge at 14.000 rpm for 5 mins (centrifuge for PCR strips).

Place the strip in a 96-well PCR microplate and centrifuge the plate upside-down at 500 rpm for 30 sec with paper tissue covering the open top to dry the EtOH.

Resuspend the DNA pellet in 20 µl H_2O by pipetting up and down 10 times. Aliquot 10 µl of each sample in two wells of a PCR microplate. Using two wells for each sample will give two separate replicates, to ensure successful sequencing. Fill all the empty wells with H_2O.

Sequence plate using a 3730 DNA Analyzer instrument (ThermoFisher, Cat. No. 3730S).

Cell culturing

This section contains general protocols for procedures used during culturing of various mammalian cell lines, including Neuro-2a, NIH/3T3 and primary mouse embryonic fibroblast (MEF) cells.

Complete culture medium preparation

All cells were cultured in complete medium:

Reagent	Percentage	Example volume
DMEM (Life Technologies, Cat. No. 31966021)	90 %	45 ml
FBS (Life Technologies, Cat. No. 10270106)	10 %	5 ml
Penicillin/streptomycin	1 %	0.5 ml
Total		50.5 ml

Thawing cells

1. Preheat culture medium to 37°C for 15 min.
2. Keep cells frozen on dry ice until you are ready to thaw them.
3. Transfer 5 ml culture medium to Falcon tube.
4. Thaw the cells in 37°C water until the ice starts to run and can be moved by inverting the tube. This is to minimize the time the thawed cells spend in the freezing medium. Remove the tube from the water and place in room temperature. The cells should be just completely thawed before transferring to culture medium.
5. Transfer 1 ml thawed cells to the 5 ml culture medium Falcon tube.
6. Centrifuge at 1000 rpm (or 300 g) for 5 min.
7. Gently discard the supernatant and carefully flick the pellet a few times to help dissolve it in the remaining solution.
8. Completely dissolve the pellet in fresh culture media by gently pipetting up and down a few times.
9. Transfer the solution to a T-flask of suitable volume and incubate at 37°C for 24h.
10. Check cell health and confluency under a microscope.
11. Change to fresh culture medium and incubate at 37°C.

Splitting/passaging cells

1. Preheat culture medium and 0.25 % trypsin/EDTA (Life Technologies, Cat. No. 25200056) solution to 37°C for 15 min.
2. Remove all medium from culture flask.
3. Gently wash the cells once with PBS.
4. Remove all the PBS and add enough trypsin to cover the entire floor of the culture flask (For T25 flasks around 1 ml is enough, for T75 3ml is enough). From this point on, work quickly to avoid excessive treatment with trypsin.
5. Incubate at 37°C for 3-5 min.
6. Check the flask under a microscope to ensure that all cells have detached. If cells are still attached, lightly thump the side of the flask against a hard surface (e.g. the side of a bench).
7. When all cells have detached, add 1:1 culture medium (e.g. for 3ml trypsin add 3ml) to neutralize the trypsin and catch all the cells in the solution.
8. Transfer the solution to a 15 ml Falcon tube and centrifuge at 1000 rpm (or 300 g) for 5 min.
9. Discard the supernatant and flick the pellet a few times to help resuspend it in the remaining solution. Resuspend the pellet in culture medium. How much medium to

add will depend on the splitting ratio. Dilute the cells as desired and transfer to fresh culture flasks.

10. Incubate at 37°C.

Seeding plates

1. Detach cells using trypsin according to the above protocol (**Splitting cells, steps 1-7**). During the incubation and/or centrifuge step, prepare a microcentrifuge tube with 10 µl dye (Trypan blue solution 0.4 %, ThermoFisher, Cat. No. C10228).
2. After centrifugation, gently remove the supernatant and resuspend the pellet in 10 ml culture medium.
3. Mix the solution gently and well, then add 1:1 cell suspension (i.e. 10 µl) to the tube with the dye.
4. Mix gently by pipetting up and down a few times, then add 10 µl to each side of a Countess Cell Counting Slide (ThermoFisher, Cat. No. C10228) by slowly releasing the solution into the slide at a 45° angle.
5. Insert the slide into Countess instrument (ThermoFisher) and measure the concentration of both sides, take the average as the concentration of the original cell suspension.
6. Dilute the cell suspension to your desired volume and concentration in a new Falcon tube. For a 24-well plate, 500 µl medium is used per well. For a 12-well plate, 1 ml medium is used per well.
7. Seed plate(s) with the diluted cell suspension. Mix gently but well to spread the cells evenly across the plate.
8. Incubate overnight at 37°C.

Freezing cells

1. Prepare freezing medium (FBS + 10% DMSO). Usually 1 ml freezing medium is used for each cryotube.
2. Count the total amount of cells (**Seeding plates, steps 1-5)** and transfer as much as needed for your cryotubes into a clean Falcon tube. 1-2 million cells per cryotube is an appropriate amount to be able to culture directly in a T75 flask after thawing.
3. Pellet cells by centrifuging at 1000 rpm (or 300 g) for 5 mins.
4. Remove the medium and resuspend the pellet in enough freezing medium to get your desired concentration.
5. Transfer the solution to cryotubes and place in a Mr. Frosty container (Thermofisher, Cat. No. 5100-0001) in -80 °C overnight. Frozen cells can then be stored in -80 °C for up to 6 months. For longer storage, place the tubes liquid nitrogen.

Transfections

All mammalian cell lines were transfected using FuGENE HD Transfection reagent (Promega, Cat. No. E2311).

Transfection timeline:

Day 0 – Seed plates
Day 1 – Transfection
Day 3 – Evaluation and photographing with EVOS instrument
Day 4 – Lysis and DNA extraction

The below table was used as a general guideline for seeding plates:

Culture vessel	Medium per well	Seeded cells per well	Total DNA per well
96-well	100 µl	8.000	100 ng
48-well	200 µl	20.000	200 ng
24-well	500 µl	40.000	500 ng
12-well	1 ml	80.000	1 µg
6-well	2 ml	200.000	2.5 µg

dsODN duplex formation:

dsODN for knock-in after Cas9 induced DSB were duplexed using single-stranded oligos. Cells in 12-well plate format were transfected with 13 pmol total dsODN per well. A list of all oligos used for dsODN formation is provided in *Appendix B*.

Add the following into a PCR tube:

Reagent	Volume
Oligo Up (100 µM)	2.6 µl
Oligo Down (100 µM)	2.6 µl
Duplex buffer (IDT, Cat. No. 11-01-03-01)	14.8 µl
Total	20 µl

Annealing condition: 5 min at 95°C, cool to RT

After annealing the concentration of dsODN is 13 µM (=pmol/µl). 1 µl dsODN is added to each well.

Transfection procedure:

Incubate Opti-MEM (Life Technologies, Cat. No. 31985070) and FuGENE at RT for 15-20 minutes before starting.

1) Dilute DNA in Opti-MEM
2) Add FuGENE, mix gently by inversion, spin down briefly.
3) Incubate 15-20 mins at RT
4) Pipette 50 µl on cells, mix gently by tilting the plate
5) Incubate at 37 °C

Reagent	Amount for one well (24-well format)	Amount for one well (12-well format)
Opti-MEM	50 µl	50 µl
Plasmid DNA	500 ng	1 µg

dsODN	-	1 µl (13 pmol)
FuGENE HD	2.25 µl	4.5 µl

Lysis and DNA extraction:

Genomic DNA was extracted 72 h post-transfection with Gentra Puregene Cell kit (Qiagen, Cat. No. 158388) (12-well plates) according to the manufacturer's instructions or QuickExtract DNA extraction solution (Lucigen, Cat. No. QE09050) (24-well plates).

QuickExtract DNA extraction:

1. Remove all medium from wells.
2. Add 125 µl of QuickExtract to the cells (for 24-well plate). Either incubate at 37 °C for 10 min or mix and scrape off the cells with a pipette tip and transfer all of the solution to PCR tubes. Remember to use a thermocycler that is accurate at 125 µl.
3. PCR program: 70 °C for 10 minutes, followed by 98 °C 10 min, cool down to 4°C.
4. Samples are ready for PCR.

QuickExtracted genomic DNA can be stored at 4 °C for at least one week or at -20 °C for several months. 1 µl of genomic DNA from QuickExtract was used per 10 µl PCR reaction.

Sanger sequencing and genotyping cell lines

The following protocol was used to sequence the gRNA on-target sites of many different cell lines, to verify correct on-target sequence. Samples were shipped to and sequenced by GENEWIZ (https://www.genewiz.com/) in Leipzig, Germany. A list of the primer sequences is provided in *Appendix B*.

1. Extracting gDNA

Extract gDNA with Lucigen QuickExtract.

1. Remove medium from cells.
2. Add 125 µl of QuickExtract to the cells (for 24-well plate). Either incubate at 37°C for 10 minutes or mix and scrape off the cells with a pipette tip and transfer all the liquid to PCR tubes. Remember to use a thermocycler that is accurate at 125 µl.
3. PCR program: 70°C for 10 minutes, followed by 98°C 10 min, cool down to 4°C
4. Samples are ready for PCR

Use 1 µl of genomic DNA from QuickExtract per 10 µl PCR reaction.

2. PCR

Amplify the region of interest with standard PCR using Phusion Flash High-Fidelity polymerase (ThermoFisher, Cat. No. F-548S)

For gP gRNA on-target site (*Pcsk9*) use PA83-84, product is 619 bp:

Reagent	Volume		Temp	Time	Cycles
2x Phusion	10 µl		98°C	10 sec	
P fwd (10µM)	1 µl		98°C	1 sec	x35
P rev (10µM)	1 µl		61°C	5 sec	
gDNA	2 µl		72°C	12 sec	
Water	6 µl		72°C	1 min	
Total	20 µL		4°C	∞	

For Nfx gRNA on-target site (*Nfx1*) use MM165-166, product is 259 bp:

Reagent	Volume		Temp	Time	Cycles
2x Phusion	10 µl		98°C	10 sec	
P fwd (5µM)	2 µl		98°C	1 sec	x35
P rev (5µM)	2 µl		60°C	5 sec	
gDNA	2 µl		72°C	10 sec	
Water	4 µl		72°C	1 min	
Total	20 µL		4°C	∞	

Verify products on agarose gel, load 3 µl of sample.

3. Purification

Purify using QIAquick PCR Purification Kit (Qiagen, Cat. No. 28106) according to the manufacturer's instructions.

4. Genewiz sample preparation

Submission to Genewiz is made in pre-mixed 1.5 ml tubes:

Template				Primer		Template + Your Primer
DNA Type	Submitted in	Concentration	Volume	Concentration	Volume*	Pre-Mixed Total Volume
Plasmids purified	1.5 ml reaction tube	30-100 ng / µl	5 µl	5 pmol/ µl (µM)	5 µl	10 µl
PCR fragments purified		10-50 ng / µl				

Add a Genewiz barcode to each tube and ship.

T7 endonuclease I assay

Alt-R genome editing detection kit (IDT, Cat. No. 1075932) was used to quickly verify editing at the on-target site in DNA extracted from Cas9 and gRNA treated cells. A list of the primer sequences is provided in *Appendix B*.

1. PCR amplification

Use Phusion Flash High-fidelity PCR Master Mix (ThermoFisher, Cat. No. F548L).

Reagent	**Amount 1x rxn (µl)**
Water	7
Phusion Flash 2x Master mix	10
Forward primer (10 uM)	1
Reverse primer (10 uM)	1
DNA (100ng/µl)	1
Total	20

Verify PCR products by running 5 µl on agarose gel. Use 5 µl for duplex formation (**step 2**).

2. DNA heteroduplex formation

2. a. Combine the following in a PCR tube:

Reagent	**Amount 1x rxn (µl)**
PCR (from **step 1**)	5
T7EI Reaction Buffer (10x)	2
Nuclease-free water	6
Total	13

2. b. PCR program:

Temperature (°C)	**Time**
95	10 min
95-85	-2 °C/sec
85-25	-0.3 °C/sec

2. c. Combine the following in a PCR tube:

Reagent	**Amount 1x rxn (µl)**
PCR (from **step 2.b**)	13
T7 endonuclease I (1 U/µl)	2
Total	15

Incubate at 37 °C for 60 min.

3. Fragment Analysis

Add 3 µl 5x Novex hi-density TBE sample buffer (ThermoFisher, Cat. No. LC6678) into samples.

Run on 10% TBE polyacrylamide gel with 230 V for 40 mins. Visualize results on a UV-table.

In vitro transcription of gRNA

sgRNAs for CIRCLE-seq were transcribed *in vitro* from gBlocks. A list of all gBlock sequences are provided in *Appendix B*. gBlocks were amplified and stored in bacteria by cloning the fragments into the pCR 2.1-TOPOTA vector using a TOPO TA Cloning Kit (Life Technologies, Cat. No. 450641).

1. Adenosine (A) adding to 3`end of gBlock

Reagent	Amount 1x rxn
gRNA gBlock	0.5 µl (5 ng)
AmpliTaq DNA Polymerase (ThermoFisher, Cat. No. N8080171)	1 µl
dNTP (2mM) (from TOPO TA Cloning Kit)	1 µl
10x PCR Buffer (+MgCl2) (ThermoFisher, Cat. No. N8080171)	1 µl
Water	6.5 µl
Total	10 µl

Incubate at 72 °C for 10 mins

2. TOPO Cloning

Reagent	Volume (µl)
gBlock	2
Salt Solution	1
Water	2
TOPO vector	1
Final volume	6

Mix and incubate for 30 mins at RT.

3. Transformation

TOPO TA Cloning Kits are optimized to work with One Shot Competent E. coli available from Life Technologies. There was no One Shot left in the lab so **DH10B** (New England Biolabs, Cat. No. C3019H) was used instead.

Preparation:

- Warm selective plates at 37 °C for 30 minutes
- Spread 50 µl X-gal + 5 µl IPTG on each LB plate and incubate at 37 °C until ready for use.
- Thaw on ice 1 vial of cells for each transformation.

Transformation:

1. Add 5 µL of the TOPO Cloning reaction from **step 2**, into a vial of competent E. coli and mix gently by flicking the tube 4-5 times.
2. Incubate on ice for 30 minutes.
3. Heat-shock the cells for 30 seconds at 42 °C without shaking.
4. Immediately transfer the tubes to ice.
5. Add 600 µL of room temperature DH10B growth medium.
6. Cap the tube tightly and shake the tube horizontally (200 rpm) at 37 °C for 1 hour.
7. Spread 200 µL from each transformation on a prewarmed selective plate and incubate overnight at 37 °C.

8. Pick 6 white or light blue colonies for analysis. Do not pick dark blue colonies. Culture them overnight in 4 ml LB medium with 0.1 % ampicillin (carbenicillin).
9. Miniprep and verify by sequencing with M13F (from TOPO TA Cloning Kit) and R. EcoRI digest 1 µg by incubating for 2 h at 37 °C.

4. EcoRI digestion for TOPO-gRNA plasmids
Digest 10 µg plasmid using EcoRI-HF (New England Biolabs, Cat. No. R3101S):

Content	**1x Volume (µl)**
Template DNA	25
Nuclease-free water	19
CutSmart buffer (NEB, Cat. No. B7204S)	5
EcoRI-HF (NEB, Cat. No. R3101S)	1
Total	50

Incubate at 37 °C for 2 hrs.

Verify digestion on agarose gel and gel extract gBlock cassette using QIAquick Gel Extraction Kit (Qiagen, Cat. No. 28706).

5. Phenol-chloroform clean-up
Use normal tubes, do not use non-stick tubes!

1) Add 1:1 volume of Phenol/Chloroform/Isoamyl alcohol (25:24:1) solution to your DNA. Vortex and centrifuge 14000 rpm 5 mins.
2) Take the supernatant, add 1:1 volume of Chloroform solution to your DNA. Vortex and centrifuge 14000 rpm 5 mins.
3) Take the supernatant. Use 3 M Sodium Acetate to bring the final concentration of sodium acetate in the solution to 0.3 M
4) Add 2 volumes of ice-cold 100% EtOH (if you want to precipitate RNA, use 2.5x).
5) Mix well by inversion and incubate at -80 °C for 1 hour. Centrifuge 14000 rpm, 30 mins at 4 °C.
6) Discard the supernatant, wash the pellet with ice-cold 70% EtOH, mix by inversion and centrifuge 14000 rpm 5 mins.
7) Remove EtOH, air-dry the tube and dissolve the pellet in RNAse-free water.

6. *In vitro* Transcription
In vitro transcription was performed using Ambion Megahortscript T7 kit (ThermoFisher, Cat. No. AM1354):

Reagent	**Amount 1x (µl)**
Template DNA (from **step 5**)	8
T7 reaction buffer	2
Mixed dNTP	8
T7 enzyme mix	2
Water	-
Total	20

Mix gently by flicking the tube, spin down. Incubate at 37 °C overnight.

Add 1 µl TURBO DNase to remove the template DNA. Incubate at 37 °C for 15 mins.

7. Purification
Purification was performed with Ambion MEGAclear Transcription Clean-Up Kit (ThermoFisher, Cat. No. AM1908).

1. Bring the RNA sample to 100 µl with Elution Solution. Mix gently but thoroughly.
2. Add 350 µl of Binding Solution Concentrate to the sample. Mix gently by pipetting.

3. Add 250 µl of 100% ethanol to the sample. Mix gently by pipetting.
4. Apply the sample to the filter, centrifuge for ~15 sec to 1 min, typically 10,000–14,000 rpm.
5. Wash with 2 × 500 µl Wash Solution. Extra centrifugation to dry filter.
6. Elute RNA from the filter with 50 µl preheated Elution Solution at 95 °C. Centrifuge for 1 min at RT (RCF 10,000–15,000 x g) to elute the RNA. To maximize RNA recovery, repeat this elution procedure with a second 50 µl aliquot of Elution Solution. Collect the eluate into the same tube.

CIRCLE-seq

Reference: Tsai et al 2018, Nature Methods, CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISCRISCRISCRISPR–Cas9 nuclease off-targets.

List of reagents:

Reagent	**Catalog Number**	**Vendor**
Gentra Puregene Tissue Kit	158689	Qiagen
Qubit dsDNA BR Assay Kit	Q32853	Thermo Fisher
Agencourt AMPure XP magnetic beads	E6260	Beckman Coulter
High throughput, "with bead", PCR-free Library Preparation Kit	KK8235	KAPA Biosystems
Enzymes and buffers		New England Biolabs
- Lamda exonuclease	M0262S (1.000 units)	New England Biolabs
- Exonuclease I (*E.coli*)	M0293S (3.000 units)	New England Biolabs
- USER enzyme	Included in NEBNext multiplex oligos	New England Biolabs
- T4 polynucleotide kinase	M0201S (500 units)	New England Biolabs
- T4 DNA Ligase	M0202S (20.000 units)	New England Biolabs
- Cas9 nuclease, *S. pyogenes*	M0386S (70 pmol)	New England Biolabs
Plasmid-Safe™ ATP-Dependent DNase	E3105K	Epicentre
NEBNext® Multiplex Oligos for Illumina® (Dual Index Primers Set 1)	E7600S	New England Biolabs
KAPA HiFi HotStart ReadyMix	KK2602	KAPA Biosystems
CIRCLE-seq hairpin adaptor (oSQT1288)	/5Phos/CGGTGGACCGATGATC /ideoxyU/ATCGGTCCACCG*T	IDT
MEGAshortscript™ Kit	AM1354	Thermo Fisher
Ambion MEGAclear kit	AM1908	Thermo Fisher

CIRCLE-seq Hairpin Adapter:

oSQT1288 /5Phos/CGGTGGACCGATGATC /ideoxyU/ ATCGGTCCACCG*T

Annealing Program: 95 °C for 5 min, -1 °C/min for 70 cycles, hold at 4 °C.

40 µl oligo (100 µM) + 60 µl nuclease-free water = 40 µM in 100 µl

Overview Notes

- Approximately 25 µg of starting DNA is required for each gRNA (or control) sample.
- 1 control sample, where Cas9:gRNA is not added to cleavage step, should be run for each unique genomic DNA source**.**
- Always bring the beads and the samples (with PCR incubation) to room temperature before the wash. Incubate both at RT for 20 min.

1. Input Quantification and Shearing

1. Quantify genomic DNA with Qubit BR dsDNA. Nanodrop overestimates your genomic DNA amount about 9x. DNA extracted from liver of WT C57BL/6N, male, 10 weeks old, was used.
2. Genomic DNA was sheared to an average length of 300 bp according to the standard operating protocol for the Covaris S2. ~ 7 µg/per covaris tube 130 µl.

130 µl sample volume - from 150 to 1,500 bp

	Vessel	**microTUBE AFA Fiber Snap-Cap (PN 520045)** 						
	Sample Volume	**130 µl**						
M220	**Holder**	M220 Holder XTU (PN 500414)						
	Insert	M220 Holder XTU Insert microTUBE 130 µl (PN 500489)						
	Temperature (°C)	20						
	Target BP (Peak)	**150**	**200**	**300**	**400**	**500**	**800**	**1,000**	**1,500**
	Peak Incident Power (W)	50	50	50	50	50	50	50	50
	Duty Factor	20%	20%	20%	10%	10%	5%	2%	2%
	Cycles per Burst	200	200	200	200	200	200	200	200
	Treatment Time (s)	330	150	65	70	50	52	90	20

3. Sheared DNA was cleaned up with 1.8X Ampure XP SPRI beads (Mix well by pipetting up-down 10 times, 10 mins incubation with beads, 10 mins incubation on the magnet, 2x 30 sec wash, 5 mins incubation for elution), and eluted in 35 µl of 1X TE buffer. 65 µl genomic DNA + 117 µl beads.

4. Pool all samples. Concentration after clean-up was measured using Qubit BR dsDNA. Fragment size was checked with Fragment Analyzer, NGS standard sensitivity gel, 12 capillary.

2. End-repair

1. For each end-repair reaction:

Component	Volume	16x (µl)
Nuclease-free H_2O	8 µl	128
KAPA End Repair Buffer (10X)	7 µl	112
KAPA End Repair Enzyme Mix	5 µl	80
Total Master Mix	**20 µl**	
Sheared genomic DNA (5 µg) (from **step 1**)	50 µl	800
Total	**70 µl**	

Prepare 5x reactions for each condition (2 Cas9 variants + 1 ctr)
End Repair Program: 20 °C for 30 min, hold at 4 °C.

2. 1.7X SPRI cleanup (120 µl of Agencourt Ampure XP beads), elute in 42 µl of 1X TE buffer, do not remove the beads.

3. A-tailing

1. For each A-tailing reaction:

Component	Volume	16x (ul)
KAPA A-tailing Buffer (10X)	5 µl	80
KAPA A-tailing Enzyme	3 µl	48
Total Master Mix	**8 µl**	
End Repaired DNA with beads (from **step 2**)	42 µl	
Total	**50 µl**	

A-tailing Program: 30 ˚C for 30 min, hold at 4 °C.

2. 1.8X SPRI cleanup (90 µl of PEG/NaCl SPRI Solution), elute in 30 µl of 1X TE buffer, do not remove the beads.

4. Adapter Ligation

1. For each ligation reaction to annealed adapter oSQT1288:

Component	Volume	16X (µl)
KAPA Ligation Buffer (5X)	10 µl	160
KAPA T4 DNA Ligase	5 µl	80
Annealed Hairpin Adapter oSQT1288 (40 µM)	5 µl	80
Total Master Mix	**20 µl**	
A-tailed DNA with beads (from **step 3**)	30 µl	
Total	**50 µl**	

Ligation Program: 20 °C for 1 h, hold at 4 °C.

2. 1X SPRI cleanup (50 µl of PEG/NaCl SPRI Solution), elute in 40 µl of 1X TE buffer.

3. Pool all samples. Qubit BR dsDNA measurement for concentration and FA run for size.

Note: Safe-stopping point, adaptor ligated samples can be stored at -20C for a few weeks.

5. Lambda Exonuclease/Exonuclease I (*E. coli*) Treatment

1. For each enzymatic treatment reaction:

Component	Volume	16x (µl)
Exonuclease I Reaction Buffer (10X)	5 µl	80
Lambda Exonuclease (5 U/µl)	4 µl	64
Exonuclease I (*E. coli*) (20 U/µl)	1 µl	16
Total Master Mix	**10 µl**	
Adapter ligated DNA (1 µg) (from **step 4**)	40 µl (13 µl DNA+ 27 µl water)	

Total	**50 µl**	

Incubation Program: 37 °C for 1 h, 75 °C for 10 min, hold at 4 °C.

2. 1.8X SPRI cleanup (90 µl of Agencourt Ampure XP beads), elute in 40 µl of 1X TE buffer, do not remove the beads.

6. USER/T4 PNK Treatment

1. For each USER/T4 PNK reaction:

Component	Volume	16x (µl)
T4 DNA Ligase Buffer (10X)	5 µl	80
USER Enzyme (1 U/µl)	3 µl	48
T4 Polynucleotide Kinase (10 U/µl)	2 µl	32
Total Master Mix	**10 µl**	
Lambda Exonuclease/Exonuclease I treated DNA with beads (from **step 5**)	40 µl	
Total	**50 µl**	

Incubation Program: 37 °C for 1 h, hold at 4 °C.

2. 1.8X SPRI cleanup (90 µl of PEG/NaCl SPRI Solution), elute in 40 µl of 1X TE buffer.

3. Pool all samples. Qubit BR dsDNA measurement for concentration and FA run for size.

7. Intramolecular Circularization

1. For each circularization reaction:

Component	Volume	10x (µl)
Nuclease-free H_2O	25 µl	250
T4 DNA Ligase Buffer (10X)	10 µl	100
T4 DNA Ligase (400 U/µl)	2 µl	20
Total Master Mix	**37 µl**	
USER/T4 PNK treated DNA (500 ng) (from **step 6**)	63 µl	
Total	**100 µl**	

Circularization Program: 16 °C for 16 h.

2. 1X SPRI cleanup (100 µl of Agencourt Ampure XP beads), elute in 38 µl of 1X TE buffer.

8. Plasmid-Safe ATP-Dependent DNase Treatment

1. For each DNase treatment reaction:

Component	Volume	11x (µl)
Plasmid-Safe Reaction Buffer (10X)	5 µl	55
ATP (25 mM)	2 µl	22

Plasmid-Safe ATP-Dependent DNase (10 U/µl)	5 µl	55
Total Master Mix	**12 µl**	
Circularized DNA (from **step 7**)	38 µl	
Total	**50 µl**	

Incubation Program: 37ºC for 1 h, 70 °C for 30 min, hold at 4 °C.

2. 1X SPRI cleanup (50 µl of Agencourt Ampure XP beads), elute in 30 µl of 1X TE buffer.

3. Pool all samples. Qubit BR dsDNA measurement for concentration and FA run for size.

9. In vitro Digestion with Cas9 and gRNA

1. For each *in vitro* digestion reaction:

Component	Volume
Cas9 Nuclease Reaction Buffer (10X)	10 µl
Cas9 Nuclease (1 µM)	9 µl
In vitro Transcribed guide RNA (300 ng/ul)	1 µl
Total Master Mix	**20 µl**

Incubate at room temperature for 10 min.

Plasmid-Safe DNase Treated DNA (250 ng) (from **step 8**)	80 µl
Total	**100 µl**

Digestion Program: 37 °C for 1 h, hold at 4 °C.

2. 1X SPRI cleanup (100 µl of Agencourt Ampure XP beads), elute in 42 µl of 1X TE buffer, do not remove the beads.

10. A-tailing

1. For each A-tailing reaction:

Component	Volume	4x (µl)
KAPA A-tailing Buffer (10X)	5 µl	20
KAPA A-tailing Enzyme	3 µl	12
Total Master Mix	**8 µl**	
Cas9/gRNA digested DNA with beads (from **step 9**)	42 µl	
Total	**50 µl**	

A-tailing Program: 30 °C for 30 min, hold at 4 °C.

2. 1.8X SPRI cleanup (90 µl of PEG/NaCl SPRI Solution), elute in 25 µl of 1X TE buffer, do not remove the beads.

11. Adapter Ligation

1. For each adapter ligation reaction:

Component	Volume	4x
KAPA Ligation Buffer (5X)	10 µl	40
KAPA T4 DNA Ligase	5 µl	20
NEBNext Adaptor for Illumina (15 µM) *	10 µl	40
Total Master Mix	**25 µl**	
A-tailed DNA with beads (from step 20)	25 µl	
Total	**50 µl**	

*NEBNext Adaptor for Illumina (#E7601A), it is included in NEBNext Multiplex Oligos, stock =15uM: /5'Phos/GATCGGAAGAGCACACGTCTGAACTCCAGTC/ideoxyU/ACACTCTTTCCTACACGACGCTCTTCCGATC*T-3

Ligation Program: 20 °C for 1 h, hold at 4 °C.

2. 1X SPRI cleanup (50 µl of PEG/NaCl SPRI Solution), elute in 47 µl of 1X TE buffer, do not remove the beads.

12. USER Enzyme Treatment

1. Add 3 µl of USER Enzyme (1 U/µl) to the adapter ligated DNA with beads (from **step 11**).
   Incubation Program: 37 °C for 30 min, hold at 4 °C.

2. 0.7X SPRI cleanup (35 µl of PEG/NaCl SPRI Solution), elute in 20 µl of 1X TE buffer.

3. Qubit BR dsDNA measurement for concentration and FA run for size.

13. PCR

1. For each PCR reaction:

Component	Volume	4x (µl)
Nuclease-free H_2O	13 µl	52
KAPA HiFi HotStart ReadyMix	25 µl	100
Total Master Mix	**38 µl**	
NEBNext i5 Primer (10 µM)	5 µl	
NEBNext i7 Primer (10 µM)	5 µl	
USER enzyme treated DNA (20 ng) (from **step 12**)	2 µl	
Total	**50 µl**	

PCR Program*: 98 °C for 45 s, 22 cycles of (98 °C for 15 s, 65 °C for 30 s, 72 °C for 30 s), 72 °C for 1 min, hold at 4 °C.

* Program is according to KAPA Library Amplification protocol.

2. 0.7X SPRI cleanup (35 µl of Agencourt Ampure XP beads), elute in 30 µl of 1X TE buffer.

3. Qubit BR dsDNA measurement for concentration and FA run for size.

Indexed PCR products were normalized and pooled. Final libraries were submitted for sequencing on an Illumina NextSeq500 by the AstraZeneca NGS/Transcriptomics team.

Gibson assembly of ObLiGaRe donor

Cloning of the AzCas9 ObLiGaRe donor vector was performed by Gibson assembly. The AzCas9-T2a-GFP cassette was amplified from the Cas9 expression vector by PCR using Phusion Flash High-Fidelity polymerase (ThermoFisher, Cat. No. F-548S), using primers that contained 30/50 bp ends homologous to the ends of the SwaI digested base vector. A list of the primers used is provided in *Appendix B*.

1. PCR amplify ORF fragment – Cas9-T2A-GFP

Reaction		**Cycling conditions**		**Cycles**
Reagent	**Volume [µl]**	**T [°C]**	**t**	
2x Phusion Flash	10	98	10s	x1
Primer Fwd (10 µM)	1	98	1s	x30-35
Primer Rev (10 µM)	1	65	5s	
Nuclease-free water	7,6	72	80s	
Plasmid (25 ng/µl)	0,4	72	1min	x1
Total	20	4	For ever	

2. Digest ObLiGaRe donor vector

Reagent	**Volume**
10x NEBuffer 3.1(NEB, Cat. No. B7203S	2 µl
SwaI (NEB, Cat. No. R0604S)	1 µl
Vector	4 µl (2 µg)
Water	13 µl

Incubate at 25 °C for 60 min.

Gel purify digested backbone and Cas9 cassette with QIAEX II Gel Extraction Kit (Qiagen, Cat. No.).

3. Gibson assembly of fragments

Reagent	**Sample**	**Pos. control**	**Neg. control**
2X Gibson Assembly Master Mix (NEB, Cat. No. E2611S)	10 µl	10 µl	10 µl
Vector (from **step 2**)	1 µl	10 µl	1 µl
Insert (from **step 1**)	4 µl	0	0
Nuclease-free water	5 µl	0	9 µl

Incubate at 50 °C for 60 min.

4. Transform into DH10-beta bacteria (New England Biolabs, Cat. No. C3019H)

1. Thaw chemically competent cells on ice.
2. Transfer 50 µl of competent cells to a falcon tube.
3. Add 2 µl of assembled product (from **step 3**) to competent cells.
4. Mix gently flicking the tube 4–5 times. Place the mixture on ice for 30 minutes. Do not mix.
5. Heat shock at 42 °C for 30 seconds. Do not mix.
6. Transfer tubes on ice for 5 minutes.
7. Add 400 µl of room temperature DH10-b growth medium to tubes.
8. Place the tube at 37 °C for 60 minutes. Shake vigorously (250 rpm) or rotate.
9. Warm selection plates (2 for each condition) to 37°C.

10. Spread 100 µl of the cells onto one of the ampicillin plates, spread 300 µl onto the other.
11. Incubate plates overnight at 37 °C.
12. Pick successfully transformed bacteria colonies with a pipette tip and add to 4 ml LB medium with 0.1% ampicillin (carbenicillin).
13. Grow at 37 °C with 500 rpm shaking overnight.
14. Use 500 µl overnight culture to inoculate 200 ml LB medium with 0.1% ampicillin (carbenicillin). The rest can be Miniprepped, quantified by nanodrop and verified by RE digestion and Sanger sequencing.

5. Miniprep and screen

Purify plasmid using QIAprep Spin Miniprep Kit (Qiagen, Cat. No. 27106) according to the manufacturer's instructions.

Digest singly with BamHI, BglI, BglII, EcoRV or others and check for correct banding and ensure all three insulators are present.

Sanger sequence assembly junction with BigDye Terminator (ThermoFisher, Cat. No. 4337456) to ensure correct insertion.

6. Scale-up

Purify the 200 ml overnight culture with an EndoFree Plasmid Maxi kit (Qiagen, Cat. No. 12362).

Appendix B – List of DNA sequences

This appendix contains a list of sequences for all DNA oligos used in experiments, including primers, knock-in dsODN inserts and artificial gene fragments (gBlocks). All oligos were ordered from Sigma-Aldrich.

Embryo amplicon NGS primers

All primers contained a 5'-sequencing adapter, compatible with Nextera XT Index Kit v2 (Illumina):

5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-[amplicon specific forward primer]-3' and
5'-GTCTCGTGGGCTCG GAGATGTGTATAAGAGACAG-[amplicon specific reverse primer]-3'

Site	Guide	Site location	Forward primer (5' to 3')	Reverse primer (5' to 3')	Product size
1	gP	chr4:106464181-106464204	TCATCAGCCAGGCCATCCTCCT	GTCCCAGGCGTCCATGTCCTTC	225
2	gP	chr11:101246644-101246667	GGATCTCGGCTTGGTGTGACCC	GCTCCGCACGTCGATTGGTACT	202
3	gP	chr9:59036131-59036154	ACAACAAAAACCCACCGCAGCG	GCGACTCCTCTGCACCTCCTC	222
4	gP	chr7:73927327-73927350	TCCAACAAACCAGCAAACCAGCA	GGAGAGGTGTAATTTTTGCCAGCTGC	209
5	gP	chr15:88958219-88958242	AGGCTACTCACCTCCCCTGCTG	CGTCTGCCTGTGGGTCTGTTCA	216
6	gP	chr5:146644326-146644349	CCAGGCCCTCGCATCACTGTAA	AGGACACTGGTGGGAGTCTTAGGA	200
7	gP	chr19:24013247-24013270	GAGAGGTGGGGAAGAGCTGGGA	GTGGCTGAAAGTGAGCTGGGGA	225
8	gP	chr10:37138620-37138643	CGGTGGCTTCTCCCTTCGCTG	TCTCTTCTGTCCTCCTCGCCAC	202
10	gP	chr2:17654859-17654882	CACCCCATCACCCGCTCAACTC	CCCTTTCCCAACCACGTGCAGA	217
12	gP	chr11:102230158-102230181	GAGGAAGGACAGGGTCGCGG	CCGCAAAGATGGAGGAGCCGT	210
1	gMH	chr3:37482962-37482985	GGTTTCCAGGAGCTGAGGGACA	CCCCCTCAGAGCAGTTTTAGGT	225
2	gMH	chr4:106457249-106457272	AGTGCAGACTCTGGAGCCCTGA	CTGTAGGCCCTGAAGTTGCCCC	218
4	gMH	chr16:33683508-33683531	CGGGTCATGGAGGCTTGGAGAC	ACCATGTGGCTTCGTGCTTACCT	201
6	gMH	chr12:111733417-111733439	GGCGCTACTCATCACTGCCTGT	CAGCTTTCCTGGAGGTGCCACA	216
8	gMH	chr5:98601716-98601739	TGTGCAGTAAAGAACAACTTTGAGG	ACTTCTGCCCAGAGGCACC	192
9	gMH	chr3:66104607-66104630	GCGCACATGTGTACTTGTGTATGGG	ATGGCCGACTTGAGAAGTGCTGG	152
14	gMH	chr16:8681106-8681129	CCCCGAGTCTCACTGACCCCTA	ACACCTTCGTTGGCCACCTCAG	222
17	gMH	chr4:93574622-93574645	ACACACCACAATGTCTCATTCCAGGA	AGGCCATTTGTACCAGCTAGATACT	219
24	gMH	chr9:118783495-118783517	GACCACAGTGCTCACATCCTCCT	TGTTTCTCTCACCGGAGCTGGC	200
29	gMH	chr8:26866365-26866388	ATTTTGGGACAGGGTTTCTCGCCT	GGTAAGGTGAAGTCAGAGGACACA	186
30	gMH	chr2:166541572-166541595	TTGGAGACAGCAGCCCAGTTGG	TGTGGTTCACATGATTGGACCCCC	207

gRNA plasmid cloning

The following oligos were used to insert the gRNA (crRNA) sequences into plasmids containing either SpCas9 or AzCas9 scaffold (tracrRNA). The first four nucleotides (highlighted in uppercase) were used to anneal the oligo duplex to the sticky ends created by BsaI digestion, to enable ligation by T4 DNA ligase.

Name	Guide	Upper strand oligo (5' to 3')	Lower strand oligo (5' to 3')
PA159-160	gP	ACCGagcagcagcggcggcaacag	AAACctgttgccgccgctgctgct
MM171-172	Nfx	ACCGcataagggatgtgtctagga	AAACtcctagacacatcccttatg

The sequences of the gRNA on-target site are:

gP – AGCAGCAGCGGCGGCAACAG

Nfx - CATAAGGGATGTGTCTAGGA

gRNA target site primers

Sequences of primers used to sequence the gP and Nfx gRNA on-target site are shown below. The genomic location (GRCm38/mm10) of the sites are:

gP - chr4:106464181-106464204

Nfx - chr4:40976492-40976514

Primers used for Sanger sequencing on-target sites and T7EI assay are shown below.

Name	Guide	Forward primer (5' to 3')	Reverse primer (5' to 3')	Product size	T_a
PA83-84	gP	GAGGCCGAAACCTGATCCTT	CTTAGAGACCACCAGACGGC	619	61
MM165-166	Nfx	TATGGGCTCCACAAGTGACA	CCAGAGGCATAACTGTGCTG	259	60

Primers used for NGS of on-target sites are shown below. Two alternative sets of primers (PA589-592) were designed to produce amplicons of shorter size, to allow for more read overlap and higher base calling quality in the case of oligo insertion.

All NGS primers contained a 5'-sequencing adapter, compatible with Nextera XT Index Kit v2 (Illumina):

5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-[amplicon specific forward primer]-3' and
5'-GTCTCGTGGGCTCG GAGATGTGTATAAGAGACAG-[amplicon specific reverse primer]-3'

Name	Guide	Forward primer (5' to 3')	Reverse primer (5' to 3')	Product size	T_a
PA307-308	gP	TCATCAGCCAGGCCATCCTCCT	GTCCCAGGCGTCCATGTCCTTC	225	72
PA561-562	Nfx	TGCTGCTGAGTTCATTCCTCA	ATGGCTGATTCTGAAGGCCA	260	63
PA563-564	Nfx	ATGCTGCTGAGTTCATTCCTCA	CCATGGCTGATTCTGAAGGC	263	63
PA589-590	Nfx	AAAACTTCTGGTCTAAATTGTGGGA	GCTGATTCTGAAGGCCATGGTT	229	61
PA591-592	Nfx	CTCAGGAGAGAAAAACTTCTGGTC	GAAGGCCATGGTTCTTGAGTGATT	231	62

Knock-in oligos

Sequences of dsODN inserts used in transfection experiments are shown below. Each pair of single-stranded DNA oligos were duplexed closely before transfection. The first batch of oligos used were either with or without a 5'-phosphate (denoted P). After initial transfection experiments, a second batch was designed. These included inserts with and without 5'-phosphates and phosphorothioate linkers (denoted *) at the ends.

First batch:

Name	Duplex ID	Strand	Sequence (5' to 3')
PA549	S	Upper	GCAAAGCACCCTATAGTGAGTCGTATTAACGT
PA550		Lower	TTGCACGTTAATACGACTCACTATAGGGTGCT
PA551	SP	Upper	P-GCAATGATCCCTATAGTGAGTCGTATTAGTGC
PA552		Lower	P-TTGCGCACTAATACGACTCACTATAGGGATCA
PA553	S	Upper	TCTAAGCACCCTATAGTGAGTCGTATTAACGT
PA554		Lower	TAGAACGTTAATACGACTCACTATAGGGTGCT
PA555	SP	Upper	P-TCTATGATCCCTATAGTGAGTCGTATTAGTGC
PA556		Lower	P-TAGAGCACTAATACGACTCACTATAGGGATCA
PA557	B	Upper	TATCCCCTATAGTGAGTCGTATTAGCAG
PA558		Lower	CTGCTAATACGACTCACTATAGGGGATA
PA559	B	Upper	P-CCAACCCTATAGTGAGTCGTATTACGTT
PA560		Lower	P-AACGTAATACGACTCACTATAGGGTTGG

Second batch:

Name	Duplex ID	Strand	Sequence (5' to 3')
PA571	B*P	Upper	P-G*T*TTAATTGAGTTGTCATATGTTAATAACGGT*A*T
PA572		Lower	P-A*T*ACCGTTATTAACATATGACAACTCAATTAA*A*C
PA573	B*	Upper	G*T*TTAATTGAGTTGTCATATGTTAATAACGGT*A*T
PA574		Lower	A*T*ACCGTTATTAACATATGACAACTCAATTAA*A*C
PA575	S*P	Upper	P-T*A*GAGTTTAATTGAGTTGTCATATGTTAATAACGGT*A*T
PA576		Lower	P-T*C*TAATACCGTTATTAACATATGACAACTCAATTAA*A*C
PA577	S*	Upper	T*A*GAGTTTAATTGAGTTGTCATATGTTAATAACGGT*A*T
PA578		Lower	T*C*TAATACCGTTATTAACATATGACAACTCAATTAA*A*C
PA579	S*P	Upper	P-T*T*GCGTTTAATTGAGTTGTCATATGTTAATAACGGT*A*T
PA580		Lower	P-G*C*AAATACCGTTATTAACATATGACAACTCAATTAA*A*C
PA581	S*	Upper	T*T*GCGTTTAATTGAGTTGTCATATGTTAATAACGGT*A*T
PA582		Lower	G*C*AAATACCGTTATTAACATATGACAACTCAATTAA*A*C
PA583	B	Upper	GTTTAATTGAGTTGTCATATGTTAATAACGGTAT
PA584		Lower	ATACCGTTATTAACATATGACAACTCAATTAAAC
PA585	S	Upper	TAGAGTTTAATTGAGTTGTCATATGTTAATAACGGTAT
PA586		Lower	TCTAATACCGTTATTAACATATGACAACTCAATTAAAC
PA587	S	Upper	TTGCGTTTAATTGAGTTGTCATATGTTAATAACGGTAT
PA588		Lower	GCAAATACCGTTATTAACATATGACAACTCAATTAAAC

gRNA gBlocks for CIRCLE-seq

The following artificial gene fragments (gBlocks) were used for TOPO cloning and *in vitro* transcription of gRNAs for CRICLE-seq. All gBlocks were ordered from Geneart Strings DNA Fragments.

Legend:

T7 promoter

gRNA (crRNA)

Scaffold (tracrRNA)

gP-SpCas9 scaffold

gaccaccttcacctacggcgtgcagtgcttcgcccgctaccccgaccacaTtaatacgactcactataGAGCAGCAGCG
GCGGCAACAGgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtc
ggtgctttttttttatagcatgct

Nfx-SpCas9 scaffold

gaccaccttcacctacggcgtgcagtgcttcgcccgctaccccgaccacaTtaatacgactcactataGCATAAGGGAT
GTGTCTAGGAgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtc
ggtgctttttttttatagcatgct

gP-AzCas9 version 1 scaffold

gaccaccttcacctacggcgtgcagtgcttcgcccgctaccccgaccacaTtaatacgactcactataGAGCAGCAGCG
GCGGCAACAGgtttcagttattcgtacgaatgaagtcactcttaaagtgagctgaaatcactaaaaattaagattgaacccg
gctactgactctgtcatccgggtttacttattttttttatagcatgct

Nfx-AzCas9 version 1 scaffold

gaccaccttcacctacggcgtgcagtgcttcgcccgctaccccgaccacaTtaatacgactcactataGCATAAGGGAT
GTGTCTAGGAgtttcagttattcgtacgaatgaagtcactcttaaagtgagctgaaatcactaaaaattaagattgaacccgg
ctactgactctgtcatccgggtttacttattttttttatagcatgct

gP-AzCas9 version 2 scaffold

gaccaccttcacctacggcgtgcagtgcttcgcccgctaccccgaccacaTtaatacgactcactataGAGCAGCAGCG
GCGGCAACAGgtttcagttattcgtgaaaacgaatgaagtcactcttaaagtgagctgaaatcactaaaaattaagattgaa
cccggctactgactctgtcatccgggtttacttattttttttatagcatgct

Nfx-AzCas9 version 2 scaffold

gaccaccttcacctacggcgtgcagtgcttcgcccgctaccccgaccacaTtaatacgactcactataGCATAAGGGAT
GTGTCTAGGAgtttcagttattcgtgaaaacgaatgaagtcactcttaaagtgagctgaaatcactaaaaattaagattgaac
ccggctactgactctgtcatccgggtttacttattttttttatagcatgct

AzCas9 cassette primers for ObLiGaRe donor vector

The following primers was used to amplify the AzCas9 cassette for cloning into the ObLiGaRe donor vector. Primers included 30 or 50 bp homology arms to enable Gibson assembly with the SwaI digested ObLiGaRe donor vector.

Legend:

30/50 bp Gibson homology arms

SwaI site

Kozak

START / STOP of transcription

Amplification of AzCas9 cassette with 30 bp homology

Forward:

5'-CTTCCTACCCTCGTAAAAAGCTTGTCCACCGCCACCATGgactacaaggacgacgacgacaaggcccct-3'

Reverse:

5'-GAGGCTGATCAGCGAGCTGCTAGATTAGGTTCActggttgaagtgcaccaggtaggccttttg-3'

Amplification of AzCas9-T2A-GFP cassette with 50 bp homology (without SwaI site homology)

Forward:

5'-
TATACCAACTTTCCGTACCACTTCCTACCCTCGTAAAAAGCTTGTCCACCGCCACCATGgactacaag
gacgacgacgacaaggc-3'

Reverse:

5'-
GGCAACTAGAAGGCACAGTCGAGGCTGATCAGCGAGCTGCTAGATTAGGTTCAtatcacttgtacagctcg
tccatgccg-3'

Amplification of AzCas9-T2A-GFP cassette with 50 bp homology (with SwaI site homology)

Forward:

5'-
TATACCAACTTTCCGTACCACTTCCTACCCTCGTAAAAAGCTTGTCCACCATTTGCCACCATGgact
acaaggacgacgacgacaaggc-3'

Reverse:

5'-
GGCAACTAGAAGGCACAGTCGAGGCTGATCAGCGAGCTGCTAGATTAGGTATTTTCAtatcacttgtac
agctcgtccatgccg-3'