Optimizing DNA double strand break repair for homologous recombination based gene therapy

zur Erlangung
des Akademischen Grades
Doktor der Naturwissenschaften

Dr. rer. Nat.

Im Fachbereich Biologie und Chemie
der Justus-Liebig-Universität Gießen

vorgelegt von

Song, Fei
aus Shanghai

Gießen, 2016

Die vorliegende Arbeit wurde in der Arbeitsgruppe Experimentale Ophthalmologie der Klinik und Poliklinik für Augenheilkunde des Fachbereichs 11 der Justus-Liebig-Universität Gießen, in der Zeit von November 2013 bis November 2016, unter der Leitung von Prof. Dr. Dr. Knut Stieger angefertigt.

Erstgutachter: Prof. Dr. Peter Friedhoff

Institute of Biochemistry

Heinrich-Buff-Ring 17

35392 Gießen

Zweitgutachter: Prof. Dr. Dr. Knut Stieger

Department of Ophthalmology

Friedrichstr.18

35392 Gießen

Erklärung

Hiermit versichere ich, die vorliegende Dissertation selbständig und ohne fremde Hilfe verfasst zu haben und keine anderen als die hier angegebenen Hilfsmittel benutzt zu haben. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht.

Gießen,

_____

(Fei Song)

Publications

Slijkerman, R. W.*, **F. Song***, G. D. Astuti*, M. A. Huynen, E. van Wijk, K. Stiegerand R. W. Collin (2015). "The pros and cons of vertebrate animal models for functional and therapeutic research on inherited retinal dystrophies." Prog Retin Eye Res 48: 137-159.

*Three first authors have contributed equally to this work.

Yanik, M., B. Muller, **F. Song**, J. Gall, F. Wagner, W. Wende, B. Lorenz and K. Stieger (2016). "In vivo genome editing as a potential treatment strategy for inherited retinal dystrophies." Prog Retin Eye Res. (In press)

F. Song, K. Stieger (2016). “Optimizing the DNA donor template for homology directed repair of double strand breaks.” (In revision)

Contents

Abbreviations

%	percent
°C	celsius
µ	micro
µl	microlitre
3’	three prime
5’	five prime
53BP1	p53 binding protein 1
A	adenine
aa	amino acid
AAP	assembly-activating protein
AAV	adeno-associated virus
ADA-SCID	adenosine deaminase deficiency-severe combined immunodeficiency
AP	alkaline phosphatase
ATM	ataxia-telangiectasia mutated
ATR	ataxia-telangiectasia RAD3-related kinase
BFP	blue fluorescent protein
BIR	break-induced replication
bp	base pair
BRCA1	breast cancer type 1
BRCA2	breast cancer type 2
BTR	BLM-TOPOIIIα-RMI1-RMI2
C	cytosine
Cas	CRISPR-associated
CDK	cyclin-dependent kinase
cDNA	complementary DNA
c-NHEJ	classical nonhomologous end joining
CoDA	Context-Dependent Assembly

CRISPR	clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
CtIP	CtBP-interacting protein
ddH_2O	double-distilled water
DDR	DNA damage response
del	deletion
dH_2O	distilled water
DMEM	Dulbecco's Modified Eagle Medium
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNA-PK	DNA-dependent protein kinase
DNA-PKcs	DNA protein kinase catalytic subunit
dNTP	deoxynucleotidetriphosphates
DMD	duchenne muscular dystrophy
DSB	double-strand break
DSBR	double-strand break repair
Fig	figure
G	guanine
GFP	green fluorescent protein
gRNA	guide RNA
HDR	homology-directed repair
HNSCC	head and neck squamous cell carcinoma
HR	homologous recombination
IDLV	integration-deficient lentiviral vector
indel	insertion and deletion
IRD	inherited retinal dystrophy
kb	kilo base
lin.	linearized
L	liter
LB	Luria broth

LCA	Leber's congenital amaurosis
LOH	loss-of heterozygosity
LPL	lipoprotein lipase
LPLD	lipoprotein lipase deficiency
LV	lentivirus
MDC1	mediator of DNA damage checkpoint 1
min	minute
ml	milliliter
MMEJ	microhomology-mediated end-joining
MRN	Mre11-Rad50-Nbs1
mRNA	messenger RNA
mut	mutated
NGS	next generation sequencing
NHEJ	nonhomologous end-joining
NLS	nuclear localization signal
Nr.	Number
OPEN	Oligomerized Pool ENgineering
pla	plasmid
PAM	protospacer adjacent motif
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
polθ	DNA polymerase theta
rAAV	recombinant adeno-associated virus
RAP80	receptor-associated protein 80
RFLP	restriction-fragment length polymorphism
RFP	red fluorescent protein
RFP168	ring finger protein 168
RFP8	ring finger protein 8
RGN	RNA-guided endonuclease
RNA	ribonucleic acid

RPA	replication protein A
RP	retinitis pigmentosa
RPGR	retinitis pigmentosa GTPase regulator
RT	room temperature
RVD	variable di-residue
SDSA	synthesis-dependent strand annealing
SIN	self-inactivating
siRNA	small silencing RNA
SOC	Super Optimal broth with Catabolite repression
SpCas9	Streptococcus pyogene Cas9
SSA	single-strand annealing
SSB	single-strand break
ssDNA	single-stranded DNA
T	thymine
TALE	transcription activator-like effector
TALEN	transcription activator-like effector nuclease
TBE	Tris-Boric acid-EDTA
TLR	traffic light reporter
TLS	translesion synthesis
tracrRNA	transactivating CRISPR RNA
VEGF	vascular endothelial growth factor
XLRP	X-linked retinitis pigmentosa
XRCC4	X-ray repair cross-complementing protein 4
ZF	zinc-finger
ZFN	zinc-finger nuclease
α	alpha
β	beta
θ	theta

List of figures

List of tables

1. Introduction

1.1 Gene Therapy

Gene therapy is defined as the transfer of nucleic acids (DNA or RNA) to patients' cells for the treatment of human diseases. It can be divided into two categories: "*in vivo*" and "*ex vivo*" gene therapy. In the first approach, the therapeutic gene is directly introduced into the patients' body (e.g. muscle, liver), while in the second approach, the patients' cells are isolated from the body, genetically modified in the laboratory and reintroduced into the patients' body.

Depending on the type of disease, gene therapy can be achieved either by addition of a functional cDNA copy as a substitute for the mutated gene to restore the normal genetic function (gene addition) or by using RNA interference to knock down the dominant negative and toxic gain-of-function gene products (gene silencing). Recently, gene targeting therapy has been brought to the forefront as a potential therapy approach for many monogenetic diseases, which aims at correcting the defective endogenous counterpart through homologous recombination based DNA double strand break repair (DSBR). Besides the correction of mutations that cause diseases, this genome-editing technology also enables the scientist to add therapeutic genes to the specific site in the genome and to precisely remove the mutated genome sequence. These innovative approaches have generated great enthusiasm and successfully moved from bench to bedside (Kaufmann, Buning et al. 2013).

1.1.1 Gene therapy in the past

In 1990, the first "*ex vivo*" gene therapy was performed by Dr. William French Anderson on two children with adenosine deaminase deficiency (ADA-SCID), a monogenetic disease leading to severe combined immunodeficiency. The T cells were isolated from these two patients and

transduced with an ADA-containing retroviral vector to express normal ADA gene. Both patients have shown increased T lymphocyte counts in the blood as well as ADA enzyme activity in one patient after the treatment (Blaese, Culver et al. 1995). Although the effects were temporary and many components remain to be perfected, it sets a new milestone in the development of gene therapy.

With this notable success, more promising results have been obtained from the clinical trials suffering from Leber's congenital amaurosis (Bainbridge, Smith et al. 2008, Maguire, Simonelli et al. 2008), haemophilia (Lheriteau, Davidoff et al. 2015), β-thalassemia (Cavazzana-Calvo, Payen et al. 2010), diabetes (Elsner, Terbish et al. 2012) and Parkinson's disease (Stoessl 2014). Between 1990 and 2013 more than 1900 clinical trials have been conducted with a varying degree of success and no major side effects reported (Kaufmann, Buning et al. 2013).

In 2004, China introduced the world's first commercial gene therapy drug (Gendicine®) into the market for the treatment of patient with head and neck squamous cell carcinoma (HNSCC) (Pearson, Jia et al. 2004, Peng 2005,Wilson 2005). Gendicine® is a recombinant human serotype 5 adenovirus genetically engineered to express the human p53 gene. The p53 gene is one of the most widely studied tumor suppressor genes, which plays a crucial role in preventing cancer formation. Because of its important role in preventing genome mutation and maintaining genome stability, the p53 gene has been described as "the guardian of the genome". After intratumoral injection of Gendicine®, the adenoviral particle delivers the therapeutic p53 gene to the cytoplasm and nucleus for the expression without integrating it in the host cells' chromosomes. In clinical trials and application, infiltration of many lymphocytes and inhibition of VEGF activity in biopsies of tumor lesions were observed. The combination of Gendicine® with radiotherapy has shown significant synergistic effects in Phase II/III clinical trials, with 64% complete regression and 29% partial regression (Peng 2005). The only side effect of Gendicine® is self-limited fever after more than five years clinical observation (Pearson, Jia et al. 2004). Gendicine® represents a remarkable medical achievement, and it opened the door for a gene therapy market in the near future.

In Europe, the first commercial gene therapeutic product Glybera® was approved by the European Commission at the end of 2012 (Miller 2012). Glybera® is an adeno-associated viral vector engineered to express lipoprotein lipase (LPL) for the treatment of familial lipoprotein lipase deficiency (LPLD). Familial LPLD represents a rare autosomal recessive disorder usually present in childhood characterized by severe hypertriglyceridemia with abdominal pain, recurrent acute pancreatitis, eruptive cutaneous xanthomata, hepatosplenomegaly, and other complications (Brunzell 1993). Familial LPLD is caused by extremely low or absent activity of LPL, the key enzyme responsible for hydrolysis of triglyceride-rich lipoproteins (Goldberg 1996). LPL interacts with circulating chylomicrons in the vascular lumen and converts triglyceride into free fatty acids (Bryant, Christopher et al. 2013). Without LPL, triglyceride cannot be depleted, leading to the accumulation of triglyceride-rich lipoproteins in the plasma (Goldberg 1996). After injection of Glybera® into the leg muscle, patients showed a long-term reduction of triglyceride levels (Gaudet, Methot et al. 2012). Although the clinical results were based on a small number of patients, the marketing authorization for Glybera® represents a landmark in the gene and cell therapy field.

Since the successes in the gene therapy area are based on the addition of a therapeutic gene for the treatment of loss of function genetic diseases, the application is very limited by other types of diseases like autosomal recessive and gain-of-function diseases. Therefore, great efforts are made in the genome editing field to cut and repair the endogens precisely though homologous recombination, which can be applied to any kind of monogenic diseases.

1.1.2 Gene transfer

Despite clinical success, the understanding of various gene transfer tools and molecular mechanisms has resulted in the development of gene therapy approaches with improved safety and therapeutic efficacy. Over the past years, genetically engineered viral and non-viral vectors are widely used for the delivery of therapeutic genes to the specific human cells.

Viral vectors, such as adenoviral vectors, adeno-associated viral (AAV) vectors, lentiviral vectors, and retroviral vectors, are the most common viral delivery vectors, which provide efficient gene transduction and effective gene expression in over 68% of gene therapy clinical trials (Santiago-Ortiz and Schaffer 2016).

AAV vectors in particular are widely used for many *in vivo* gene transfer applications to both dividing and non-dividing cell populations with low host immune response, high transferring ability, long-term gene expression, and low toxicity (Kotterman and Schaffer 2014). As a result of these properties, AAV is becoming a promising approach to treat a variety of diseases and cancers, including hemophilia B, LPLD, Alzheimer's disease, Parkinson's disease, inherited retinal dystrophies (IRD), and liver cancer (Luo, Luo et al. 2015).

AAV is a linear single-stranded DNA (ssDNA) parvovirus. The wild-type AAV genome, which is about 4.7 kb long, comprises two open reading frames encoding Rep and Cap protein flanked by two hairpin palindromic repeat sequences termed inverted terminal repeats (ITRs) (Santiago-Ortiz and Schaffer 2016). The Rep gene codes for non-structural proteins expressed via alternative promoters, which are involved in viral replication, transcription, packaging, and genomic integration. The Cap gene encodes the structural proteins VP1, VP2, VP3 that assemble to form the viral capsid (Deyle and Russell 2009). In the absence of a helper virus, the AAV genomes can establish latency and persist in the host as episomes or in some cases integrate into the host genome on the long arm of chromosome 19 in a site-specific manner. In the presence of a helper virus such as adenovirus or herpesvirus, the AAV can enter its replication cycle and undergoes productive infection (Deyle and Russell 2009). Infection is initiated by low-affinity binding to glycans followed by cell surface receptor-mediated endocytosis. Endosomal escape, endosomal trafficking, and viral capsid uncoating occurs and the single stranded DNA undergoes double strand synthesis, which is capable of transcription and gene expression(Lisowski, Tay et al. 2015).

Recombinant adeno-associated viruses (rAAVs) are generated by replacing the viral ORFs with transgene expression cassettes containing the gene of interest. Rep and Cap genes, as well as

helper genes required for AAV replication are provided *in trans* **(Fig. 1)**. Either single-stranded DNA genomes of 4.7 kb or self-complementary DNA genomes of approximately 2.2 kb can be packed into AAV vectors (Gray and Zolotukhin 2011). For large therapeutic genes (> 4.7 kb), different strategies have been developed to expand the packaging capacity imposed by the viral genome. These include minimizing the expression elements in the expression cassette, truncating the gene itself, or using *trans-splicing* and overlapping vectors based on post-transduction concatemerization of the transgene (Chira, Jackson et al. 2015).

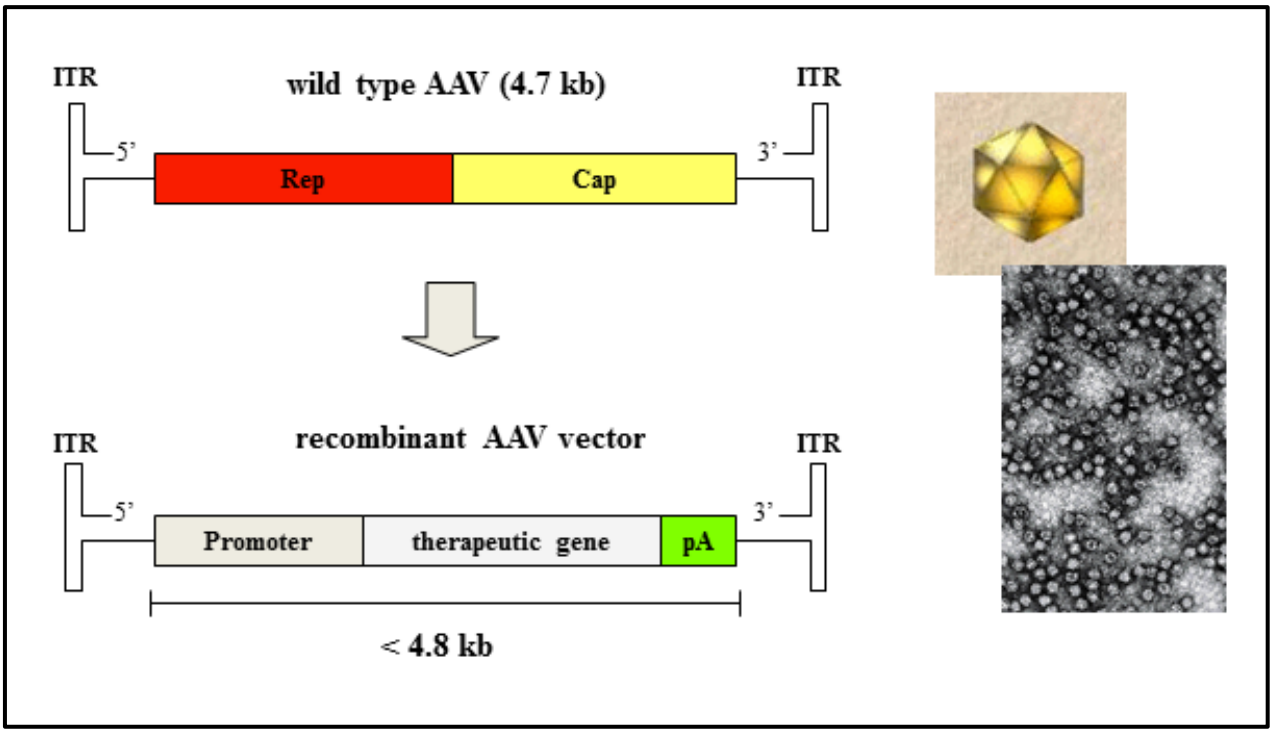

Fig. 1: Generation of recombinant AAV vectors. Recombinant AAV vectors are generated by replacing the viral ORFs with transgene expression cassettes containing the promoter, therapeutic gene and pA sequence. At the left side of the picture is the hexagonal three-dimensional structure of the viral capsid and electron microscopy image of the individual 21nm large virus (Stieger and Lorenz 2008).

However, their applications are limited to the inherent proprieties of the target gene. A hybrid dual vector system has been developed as a potential solution to this problem by inserting a

highly recombinogenic *alkaline phosphatase* (AP) sequence in the *trans-splicing* vector, which allows the AP sequence-mediated transgene reconstitution through homologous recombination (Ghosh, Yue et al. 2008). These recent advances may offer a strong potential for large gene reconstitution, however, further understanding of vector-host-interaction and endogenous mechanisms are required for achieving successful therapeutic application.

Another important type of vector for gene therapy is derived from lentiviruses (LV), which have been used particularly for the treatment of central nervous system disorders. Lentiviruses are able to transduce both dividing and non-dividing cells with high transduction efficacy, long-term stable expression of the transgene, and low immunogenicity (Chira, Jackson et al. 2015). Typically, vector particles are produced by cotransfection of the lentiviral vector plasmid and three helper constructs (pMDL, pRev and pVSVG) in the packaging cell. The vector plasmid contains the vector genome and the transgene, while the helper constructs encode the proteins that are essential for the viral life cycle (Tiscornia, Singer et al. 2006). Compared to AAVs, LV can accommodate larger transgenes up to 10 kb, which broadens the applicability of LVs for gene therapy application (Matrai, Chuah et al. 2010) **(Fig. 2)**.

Upon transduction of the target cells, the virion enters the host cells by binding to the cell surface receptors and co-receptors, leading to endocytosis or direct fusion with the cell membrane. After entry, the internal core is released and the RNA transgene is reverse transcribed to the cDNA. The viral core containing the cDNA is then transported to the nucleus and the cDNA is integrated into the host cell chromosome, resulting in persistent expression of the transgene (Tang, Kuhen et al. 1999). However, activation of proto-oncogenes at the site of genomic integration suggests the carcinogenic effect of LVs in the host genome. To overcome this limitation, self-inactivating (SIN) vectors and integration-deficient lentiviral vectors (IDLV) have been developed by deleting the U3 region of the 3'LTR and mutational inactivation of the integrase protein, respectively (Matrai, Chuah et al. 2010). These recent achievements highlight the improved safety of LVs with important implication for further clinical purposes.

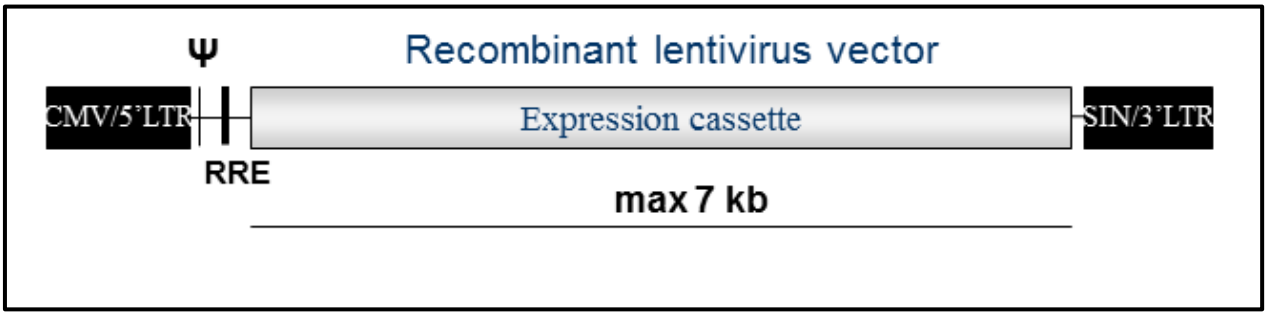

Fig. 2: Recombinant lentivirus vectors. The viral ORFs of the wild type Lentiviruses are replaced by the expression cassette of the therapeutic gene, which is maximal 7 kb long (Stieger and Lorenz 2014). LTR: Long terminal repeat; SIN: Self inactivating LTR; RRE: Rev Response Element; Ψ: Pack sequence psi.

Despite accumulating data on improved viral vectors, non-viral vectors, such as naked plasmids, nanoparticles, cationic liposomes and cationic polymers have also shown their advantages in the gene therapy field. In contrast to viral-vectors, non-viral vectors are typically easy to synthesize, and the immunogenicity is lower. Non-viral vectors are also capable of delivering large therapeutic genes and synthetic expression cassettes like siRNA. The limitations of non-viral vectors are related to extracellular stability, internalization, intracellular trafficking, nuclear entry, and the sustainability of expression of the transgene. Several strategies have been developed to overcome these limitations by using carrier molecules, targeting ligands, endosomal disruptive agents, and nuclear localization signal (Ramamoorth and Narvekar 2015).

Besides these recent improvements in gene delivery vectors and a large amount of clinical successful events, the development of programmable nucleases and the understanding of nuclease functions opened the new era of genome editing field.

1.2 Highly specific nucleases

In recent years, several programmable nucleases like meganucleases, zinc-finger nucleases (ZNFs), transcription activator-like effector nucleases (TALENs) and RNA-guided

endonucleases (RGNs) have become powerful tools for precise and efficient genome engineering. These engineered nucleases can create site-specific DNA double-strand breaks, and the induced double-strand breaks can be repaired through different DNA repair pathways including homologous recombination, non-homologous end joining, and microhomology-mediated end joining **(Fig. 3)**. To induce site-specific DNA DSB in the human genome of 3 x 10^9bp, the recognition site of the nucleases should be 16-18 bp in length (Chandrasegaran and Carroll 2016).

To date, ZNFs and TALENs have been used in more than 40 different organisms and cell types, which have shown their successes in genome editing area (Chandrasegaran and Carroll 2016). A more recent genome editing tool is the CRISPR-Cas9 system, which has been described as an adaptive defense mechanism in bacteria and archaea. In the past three years, CRISPR-Cas systems have been used in a wide range of organisms, including *Drosophila melanogaster* (Gratz, Cummings et al. 2013), *Caenorhabditis elegans* (Friedland, Tzur et al. 2013), *Saccharomyces cerevisiae* (DiCarlo, Norville et al. 2013), zebrafish (Chang, Sun et al. 2013), mice (Wang, Yang et al. 2013), rat (Li, Qiu et al. 2013, Li, Teng et al. 2013), plants (Jiang, Zhou et al. 2013), monkeys (Niu, Shen et al. 2014) and human embryos (Kang, He et al. 2016).

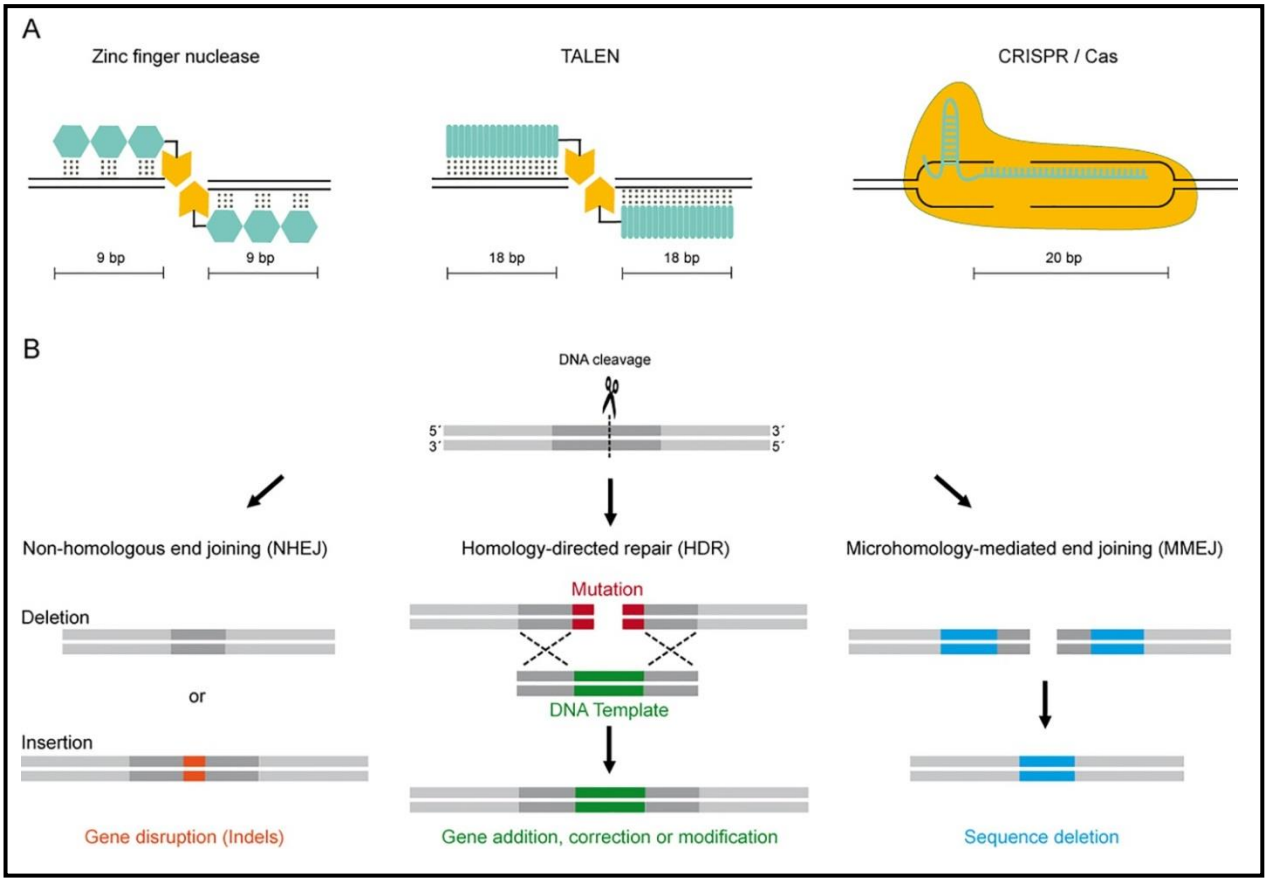

Fig. 3: Description of the different programmable nucleases and three major repair pathways. (A) Schematic drawing of ZNF, TALEN, and CRISPR-Cas. The DNA binding domains are represented in turquoise, the cleavage domains are in yellow. **(B)** Three major repair pathways occur after the introduction of the DNA double strand breaks, including NHEJ, HDR, and MMEJ (Yanik, Muller et al. 2016).

1.2.1 Zinc finger nucleases

ZNFs were the first artificial targetable nucleases, which can be customized to cleave any given sequence in the genome (Kim and Berg 1996). ZFNs consist of a programmable DNA-binding domain of zinc finger proteins (ZFPs) and a non-specific DNA cleavage domain of type II restriction endonuclease FokI (Kim, Cha et al. 1996). Because the binding domain of the ZFPs can be easily manipulated, they have become powerful tools for genome engineering and have

been used to site-specifically modify the genomes in various organisms, including frogs, insects, fish, plants, mice, rats and cultured human cells (Wu, Kandavelou et al. 2007). Furthermore, ZFNs are the only genome editing tools that have been tested in clinical trials for the treatment of Hemophilia B, Mucopolysaccharidosis I and HIV infections (Yanik, Muller et al. 2016).

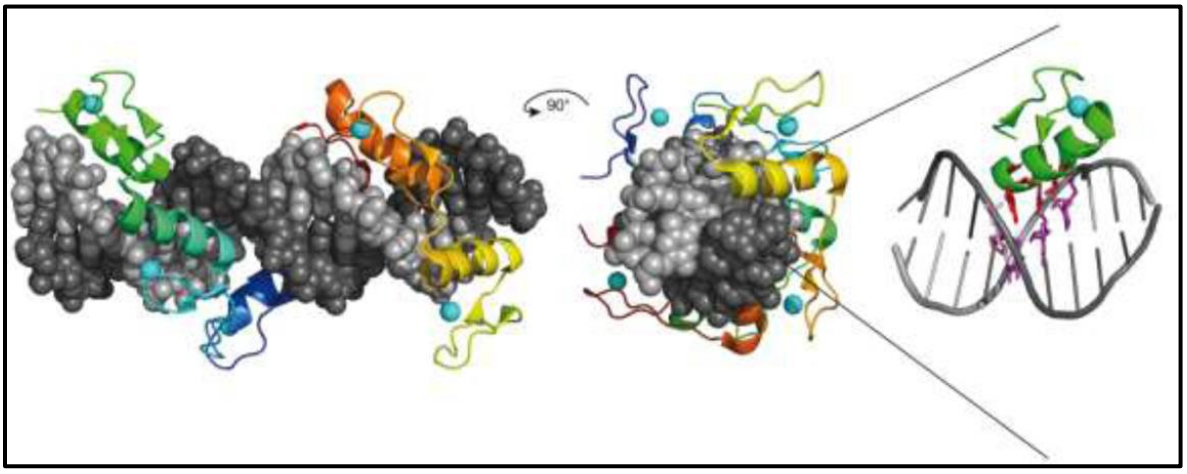

Fig. 4: Schematic representation of DNA binding by zinc finger protein. Each zinc finger usually recognizes 3 base pairs DNA through four contact amino acids at positions 1, 2, 3, 6 of each helix (shown in purple) (Pingoud, Wilson et al. 2014).

The DNA-binding domain of ZFNs contains 3 to 6 Cys_2-His_2 zinc fingers. Each zinc finger (ZF) is composed of approximately 30 amino acid residues in a unique ββα configuration, which is stabilized by a zinc atom (Pavletich and Pabo 1991). An individual zinc finger usually recognizes 3 base pair DNA sequences by inserting an α-helix into the major groove of the DNA double helix (Pavletich and Pabo 1991) **(Fig. 4)**. The binding specificity of the ZFP has a direct influence on the cleavage specificity, which can be achieved by adding more ZF motifs to the ZFPs, but the recognition of DNA by individual ZFs appears to be dependent of the neighboring modules (Urnov, Miller et al. 2005, Urnov, Rebar et al. 2010). To construct multi-finger arrays and to improve the binding specificity, different strategies have been developed, including OPEN (Oligomerized Pool ENgineering) and CoDA (Context-Dependent Assembly) (Maeder, Thibodeau-Beganny et al. 2008, Sander, Dahlborg et al. 2011).

To produce a DSB, the nuclease domain of the FokI endonuclease must be dimerized for the formation of the active center (Bitinaite, Wah et al. 1998, Wah, Bitinaite et al. 1998). Therefore, a pair of ZFNs binding to adjacent, oppositely oriented sites on the DNA is required for the induction of the DSB (Smith, Bibikova et al. 2000). Paired binding sites of ZFNs doubled the size of the target sequence recognition and increased the specificity of ZFNs. Normally, a pair of 3- or 4- finger ZFN monomers has an 18- or 24- bp recognition site in a tail-to tail orientation, which is long enough to specify a unique genomic sequence in mammals and plants (Wu, Kandavelou et al. 2007). However, due to the unspecific DNA cleavage domain and unexpected binding to nonspecific DNA sequences, significant off-target cleavage effects have been observed besides the on-target cleavage, indicating that still great efforts should be made to improve the specificity of the ZFNs (Gabriel, Lombardo et al. 2011, Pattanayak, Ramirez et al. 2011).

1.2.2 TALENs (transcription activator-like effector nucleases)

TALENs (transcription activator-like effector nucleases) are another type of highly specific nucleases, which can be designed to cleave any given DNA sequences for targeted modification of endogenous genes. The general modular structure of TALENs is similar to that of ZFNs, which is based on a specific DNA binding domain of bacterial TALEs fused to the unspecific DNA cleavage domain.

The DNA binding domains TALEs (transcription activator-like effectors) are transcriptional activators that are derived from the bacterial plant pathogen *Xanthomonas* (Bonas, Stall et al. 1989). TALEs contain N- and C- termini for localization and a central repeat domain for specific DNA binding (Boch and Bonas 2010). The central repeat domain of TALE proteins is composed of 5 to over 30 tandem repeats with an average of 17.5 (Wei, Liu et al. 2013). Each repeat contains 33-35 amino acid residues and recognizes one base pair in the target DNA via unique repeat variable di-residues (RVD) at position 12 and 13 (amino acid residue NI recognizes A, HD recognizes C, NG or HG recognizes T, and NN recognizes G or A), which determines the nucleotide binding specificity of each repeat (Deng, Yan et al. 2012, Mak, Bradley et al. 2012)

(Fig. 5). Since there is no considerable context-dependent interaction between the neighboring repeats, TALENs have much more advantages compared to ZFNs (Yanik, Muller et al. 2016).

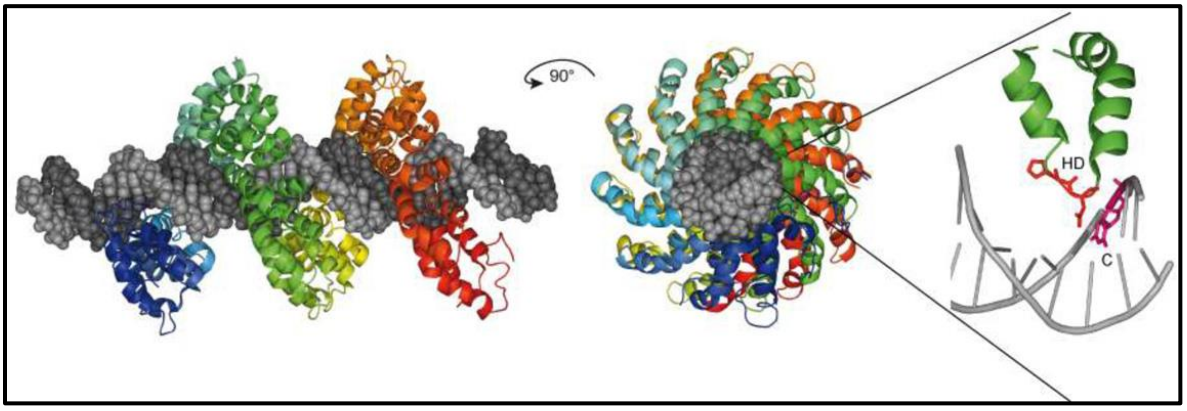

Fig. 5: Schematic representation of DNA binding by TAL effector. The DNA binding domain of TALEs consists of a series of repeats and each repeat recognizes one base pair in the target DNA via repeat variable di-residues (RVD) at position 12 and 13. The RVD (HD) is shown in red (Pingoud, Wilson et al. 2014).

By fusing the nuclease cleavage domain such as FokI and PvuII with an artificial TALE binding domain, TALENs have been used in a wide range of model organisms and cell types, including flies, frogs, fish, rats, mice, human somatic cells, and human cells (Miller, Tan et al. 2011, Wei, Liu et al. 2013, Yanik, Muller et al. 2016). In 2015, the first-in-man application of TALEN engineered universal CAR19 T cells took place and will now be tested in clinical trials (Yanik, Muller et al. 2016).

1.2.3 CRISPR-Cas system

Recently, the development of CRISPR-Cas (clustered regularly interspaced short palindromic repeats-CRISPR-associated proteins) systems is revolutionizing the field of genome editing, enabling the scientists to manipulate the genomes for therapeutic application with relative ease.

The CRISPR-Cas systems are prokaryotic immune systems, which provide adaptive immunity against invading phages and foreign nucleic acids. CRISPR-Cas systems are present in roughly half of all sequenced bacterial genomes and almost all of sequenced archaeal genomes (Barrangou, Fremaux et al. 2007, Wright, Nunez et al. 2016).

In these species, CRISPR arrays contain a series of short, palindromic DNA repeats ranging from 21 to 48 bp, interspaced by 26 to 72 bp variable spacer sequences derived from invading pathogens (Bondy-Denomy and Davidson 2014). The CRISPR array is usually located adjacent to a cluster of CRISPR-associated (cas) genes and preceded by an AT-rich leader sequence (Amitai and Sorek 2016). Depending on the Cas genes and the proteins they encode, three major types of CRISPR-Cas systems have been identified, namely Type I, Type II and Type III, and these can be further divided into several subtypes, given their structural and functional diversity (Amitai and Sorek 2016). The key protein of Type I systems is Cas3, while the signature protein of Type III systems is Cas10. The most widely used CRISPR systems are the Type II CRISPR-Cas9 systems from *Streptococcus pyogenes*, which are signified by Cas9 protein (Wright, Nunez et al. 2016).

Overall, CRISPR immunity can be divided into three stages: adaption, CRISPR RNA (crRNA) biogenesis, and interference (van der Oost, Westra et al. 2014, Makarova, Wolf et al. 2015). In the adaption stage, foreign DNA is identified and integrated into the CRISPR array as a new spacer. During the crRNA biogenesis stage, the CRISPR array is transcribed and processed into small crRNAs that each contains a single spacer flanked by CRISPR repeat sequences. These mature crRNAs are subsequently combined with Cas proteins for the formation of the active Cas-crRNA complex. In the interference stage, the Cas-crRNA complex recognizes foreign nucleic acid sequence complementarity to the crRNA sequence, which leads to the successful cleavage and degradation of the DNA and RNA molecules **(Fig. 6)**.

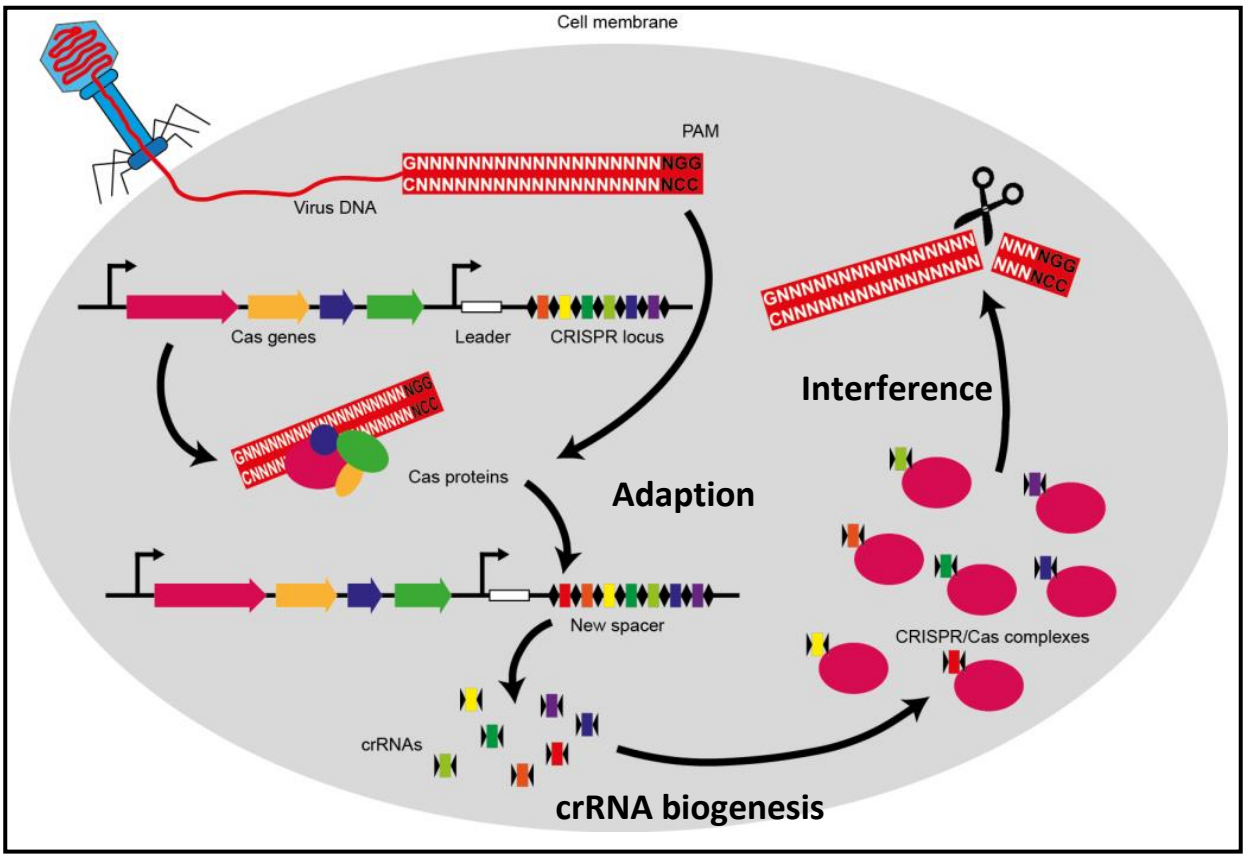

Fig. 6: Three stages of CRISPR immunity: adaption, crRNA biogenesis, and interference. During the adaption stage, foreign DNA is identified and integrated into the CRISPR array as a new spacer. In the crRNA biogenesis stage, the CRISPR array is transcribed and processed into small crRNA, which combines with Cas protein and forms the active Cas-crRNA complex. The Cas-crRNA complex recognizes and cleaves the foreign DNA 3 bp upstream of the PAM sequence.

The engineered Type II CRISPR-Cas9 system consists of the Cas9 DNA endonuclease and a chimeric single guide RNA (gRNA). The gRNA contains a 20 nt programmable CRISPR RNA (crRNA) and a trans-activating crRNA (tracrRNA), which forms a double-stranded RNA structure and binds to Cas9that form a complex for cleavage (Jinek, Chylinski et al. 2012, Hsu, Lander et al. 2014). Next to the target sequence, a 2-5 nucleotide motif is required for the target recognition; named protospacer adjacent motif (PAM) that usually consists of a 5'-NGG-3' trinucleotide, in which N can be any nucleotide (Horvath and Barrangou 2010). Upon recognition

of gRNA and target DNA sequence, the Cas9-gRNA-target DNA ternary complex initiates the subsequent cleavage of the target strand by the HNH nuclease domain, and of the non-target strand by the RuvC domain **(Fig. 7)** (Amitai and Sorek 2016).

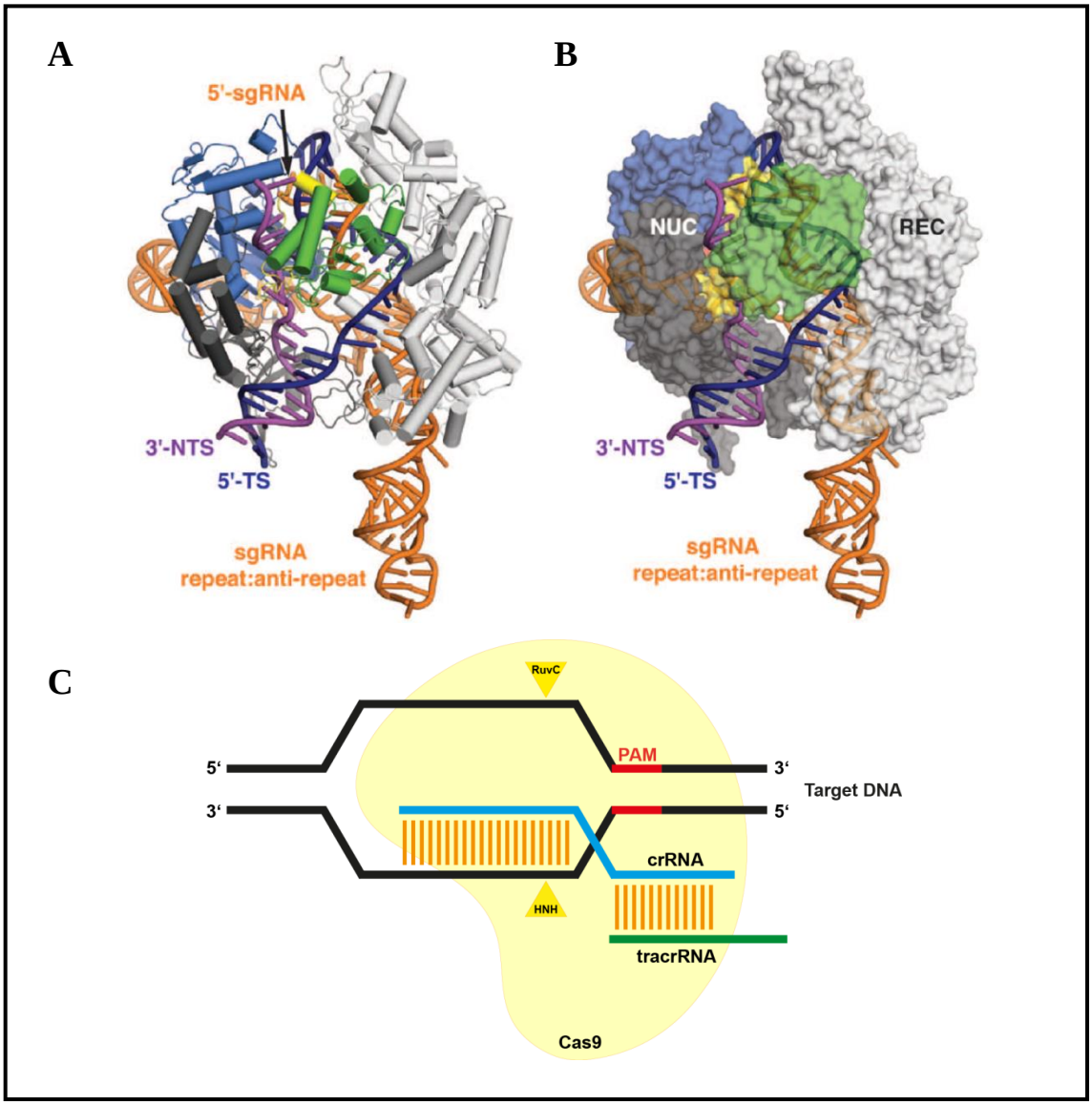

Fig. 7: CRISPR-Cas9 mediated DNA double strand break. (A and B) Crystal structure of Cas9-sgRNA-dsDNA ternary complex. The target DNA strand and nontarget strand are colored dark blue and purple, respectively. sgRNA is shown in orange (Amitai and Sorek 2016) **(C)** Upon the recognition of 20 bp target DNA next to the PAM sequence, DNA double strand break is mediated by the two activate centers of the Cas9 protein, RuvC and HNH. TS, target strand; NTS, nontarget strand.

Beyond the introduction of the DNA breaks, CRISPR technologies have also been used in different ways, such as regulating gene expression, modifying epigenomes, and dynamic imaging of chromatin. Specifically, the CRISPR-associated catalytically inactive dCas9 protein has been fused to transcription repressor or activator domains for the regulation of gene expression in human and yeast cells (Gilbert, Larson et al. 2013, Larson, Gilbert et al. 2013,Gilbert, Horlbeck et al. 2014). By using eGFP-tagged dCas9 protein and sequence-specific gRNA, dCas9 chimeras enable the imaging of DNA and visualization of chromatin organization and dynamics in living human cells (Chen, Gilbert et al. 2013). Likewise, the CRISPRainbow technique based on dCas9 combined with fluorescence-labelled gRNA has demonstrated simultaneous imaging of up to six distinct chromosomal loci in living cells (Ma, Tu et al. 2016). Recently, CRISPR-Cas9-based acetyltransferases and demethylases enable the epigenetic regulation and provide a new tool for manipulating gene expression (Hilton, D'Ippolito et al. 2015, Pham, Kearns et al. 2016).

1.3 DNA double strand break repair

1.3.1 DNA damage responses (DDR)

In human cells, DNA damage takes place at a rate of 10,000 to 1,000,000 molecular lesions per cell per day. It can be caused by exogenous agents and endogenous cellular processes, such as ionizing radiation (IR), ultraviolet light (UV), chemical agents, and replication errors (Hoeijmakers 2001, Hoeijmakers 2001). One of the most dangerous types of DNA damage is a double-strand break (DSB) that can result in the introduction of gene mutations, chromosome rearrangement, and cell death (Khanna, Lavin et al. 2001). DNA double strand breaks can also be induced artificially by using highly specific nucleases and through the addition of template DNA to trigger the desired repair outcomes. Efficient and accurate DNA repair is crucial for the maintenance of genomic stability and prevention of tumor formation. Designed DNA cut and repair can be used as a potential therapeutic approach to repair the disease causing mutations. In mammalian cells, DNA double-strand breaks (DSBs) can be repaired by different pathways,

homologous recombination (HR), classical nonhomologous end joining (c-NHEJ), microhomology-mediated end joining (MMEJ), and single-strand annealing (SSA) (Helleday, Petermann et al. 2008). In the cell division cycle, multiple cyclin-dependent kinases (CDKs) are periodically activated and play a central role in DNA repair pathway choices. C-NHEJ can occur in any phase of the cell cycle but is dominant in G0/G1 and G2, whereas HR usually takes place in S and G2 cell phases because it uses sister-chromatid sequences as the template for repair (Sancar, Lindsey-Boltz et al. 2004).

In addition to these repair pathways, DNA damage response (DDR) also plays a key role in combating threats posed by DNA damage. It is a signal transduction pathway that enables the cell to detect DNA lesions, propagate DNA damage signals, and promote their repair. DDR pathway is mediated by the activation of ataxia-telangiectasia mutated (ATM), ataxia-telangiectasia RAD3-related kinase (ATR), and DNA-dependent protein kinase (DNA-PK). ATM is primarily activated by double-stranded DNA breaks (DSBs), whereas ATR responds to RPA-coated single-stranded DNA (ssDNA) region.

The key regulator of ATM activation is the Mre11-Rad50-Nbs1 (MRN) complex, which functions as a DSB sensor and is required to recruit DDR downstream proteins. DNA-PK promotes DSB religation and is involved in the non-homologous end joining pathway of DNA repair (Goldstein and Kastan 2015). Once the activated ATM and ATR kinases are at the DSB, they phosphorylate a number of substrates such as H2AX, NBS1, CHK1, BRCA1, p53, and CHK2. The phosphorylated form of H2AX (known as γH2AX), is recognized by the mediator of DNA damage checkpoint 1 (MDC1), which then recruits ring finger protein 8 (RNF8), an E3 ubiquitin ligase. RNF8 promotes another E3 ligase ring finger protein 168 (RNF168) to ubiquitinate H2A-type histones, leading to the recruitment of the p53 binding protein 1 (53BP1) and receptor-associated protein 80 (RAP80) to DSB sites (Bohgaki, Bohgaki et al. 2013).

After the DNA damage responses, different repair key proteins affect the decision of the repair pathway choices with collaboration and competition (Kass and Jasin 2010). Understanding of the

repair mechanisms helps us to inhibit the error-prone repair pathway and bias the repair outcomes toward HDR.

1.3.2 Homologous recombination (HR)

HR is a central pathway for accurate DNA double strand break repair (DSBR), which is described as an error-free repair mechanism. HR uses an undamaged homologous sequence as a donor template for repair and requires RAD51-mediated strand invasion. HR is initiated by resection of DNA ends at the DSB site. In most cases, DNA end resection in eukaryotes is a two-step process (Mimitou and Symington 2008, Zhu, Chung et al. 2008,Ceccaldi, Rondinelli et al. 2016). In the initial phase of the end resection, a small number of base pairs (fewer than 20 bp in mammalian cells, 100-200 bp in yeast) are processed by MRE11-RAD50-NBS1 (MRN) complex and CtBP-interacting protein (CtIP) (Zhu, Chung et al. 2008, Truong, Li et al. 2014).

Following the initial DNA processing, the extension resection (which is known as 5'-3' resection) is mediated by helicases and nucleases (i.e., CtIP, EXO1, DNA2, BLM, WRN) to generate a long stretch of 3' single-strand DNA (ssDNA) for strand invasion (Sturzenegger, Burdova et al. 2014, Ceccaldi, Rondinelli et al. 2016). During this process, end resection is promoted by cyclin-dependent kinase (CDK)-dependent phosphorylation of multiple substrates, such as DNA2, CtIP, and EXO1 (Yun and Hiom 2009, Chen, Niu et al. 2011,Tomimatsu, Mukherjee et al. 2014). Next, the resected DNA is coated by ssDNA-binding protein replication protein A (RPA) to minimize the formation of secondary structures. RPA is then displaced and the Rad51 is loaded onto the ssDNA to form a nucleoprotein filament, a step that is mediated by BRCA2 (breast cancer type 2) and RAD52 (Renkawitz, Lademann et al. 2014). After RAD51 formation, homology search and DNA strand invasion takes place leading to D loop formation between the broken DNA and the intact homologous donor sequence (Sancar, Lindsey-Boltz et al. 2004).

More than three different pathways are proposed after the D-loop intermediate. In the classic double-strand break repair (DSBR) model, the 3' end in the D loop is extended by repair

synthesis and the second DSB end aligns with the extended D‑loop to form a double holliday junction (DHJ). Crossover and non-crossover overcomes are produced by resolvases or combined helicase/topoisomerase, such as GEN1, MUS81-EME1 complex, and BLM-TOPOIIIα-RMI1-RMI2 (BTR) complex (Matos and West 2014). According to the synthesis-dependent strand annealing (SDSA) model, the nascent strand from the D-loop anneals to the 3' end of the broken chromosome and the single-stranded gaps are filled in by DNA synthesis and ligation. This type of repair results in non-crossover products (Rodgers and McVey 2016). In break-induced replication (BIR), only one end of the DSB aligns homology with the template and the D-loop is assembled into the replication fork **(Fig. 8)**. Replication of the entire homologous template arm results in a large-scale loss-of heterozygosity (LOH) (Pardo, Gomez-Gonzalez et al. 2009).

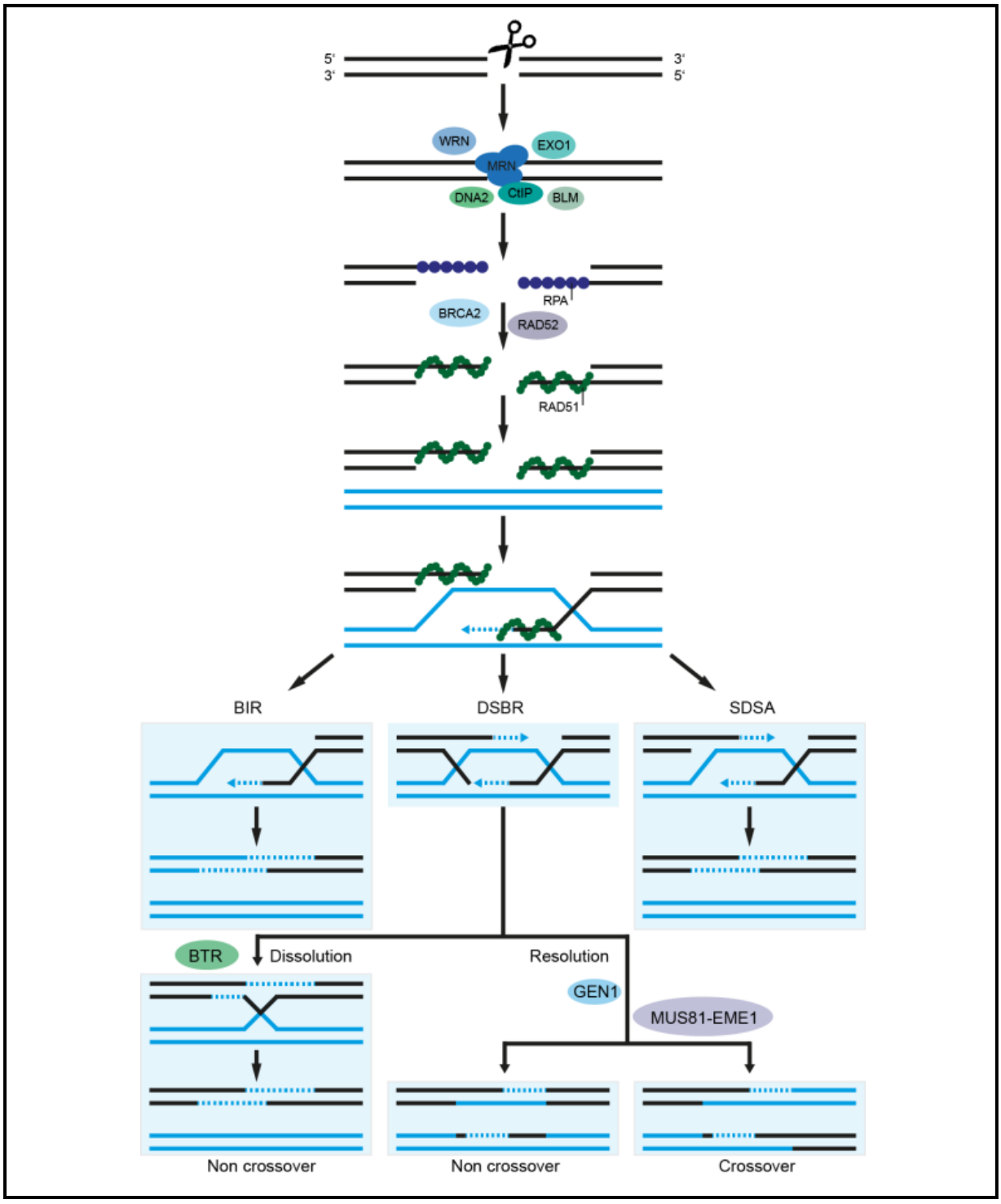

Fig. 8: Homology-directed DNA repair pathway. HDR is initiated by DNA end resection at the DSB site, followed by the binding of RPA and Rad51. In the presence of DNA donor template, precise DNA repair takes place resulted in crossover or non crossover products.

1.3.3 Non homologous end joining (NHEJ)

NHEJ has been considered as error-prone repair of DSBs, which is characterized by re-ligating of two broken ends independently of sequence homology. NHEJ is a predominant DNA double strand repair pathway, which is always associated with the introduction of small insertions and deletions (indels) at the break site. Non homologous end joining is initiated by the binding of the KU70/80 heterodimer, followed by recruitment and activation of the DNA protein kinase catalytic subunit (DNA-PKcs). Binding of the Ku heterodimer protects the broken ends from extensive resection and inhibits their degradation. DNA-PKcs undergoes autophosphorylation and phosphorylates other NHEJ downstream proteins, such as X-ray repair cross-complementing protein 4 (XRCC4) and Artemis. DNA ends are subsequently re-ligated by the XRCC4-ligase IV-XLF complex (Hoeijmakers 2001, Helleday, Petermann et al. 2008) **(Fig. 9)**.

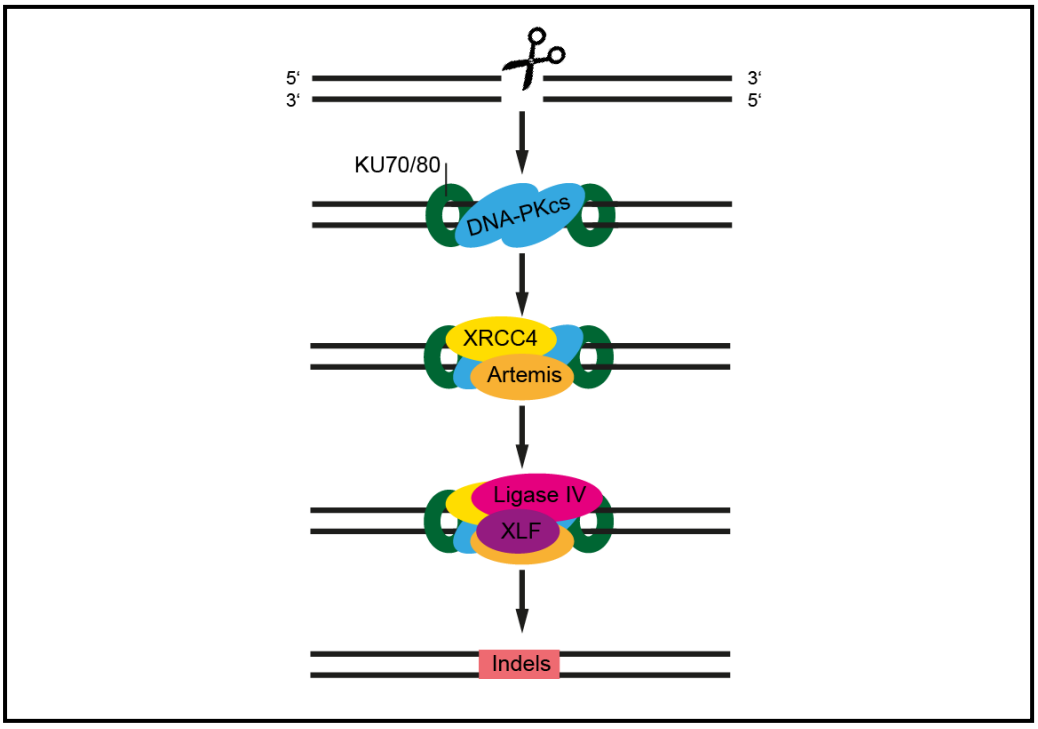

Fig. 9: Non homologous end joining DNA repair pathway. After the DNA DSBs induced by specific nucleases, KU70/80, DNA-PKcs, XRCC4, and Artemis bind at the break site. The ends are finally re-ligated by XRCC4-ligase IV-XLF complex, which always results indels.

1.3.4 Microhomology-mediated end joining (MMEJ)

In the absence of c-NHEJ factors such as Ku70, Ku80, or DNA ligase IV, yeast and mammalian cells are still able to repair DSBs via an alternative form of DSBR, termed microhomology-mediated end joining (MMEJ) (Deriano and Roth 2013). MMEJ is an error-prone repair mechanism that always generates small deletions flanking the DSBs. MMEJ is also associated with chromosomal translocations and telomere fusions, thereby resulting in harmful consequences on genomic stability (McVey and Lee 2008).

In mammalian cells, both MMEJ and HR share the common initial end resection step mediated by MRN complex and CtIP in a CDK-dependent manner, which reveals 5-25 base pair (bp), single-strand microhomologous sequences for further sequence alignment (McVey and Lee 2008). Repair is completed by annealing of microhomologies, cleavage of 3' flaps, fill-in DNA synthesis, and ligation. Although the MMEJ repair mechanism is still less characterized, numerous studies highlight critical roles for XRCC1/DNA ligase III complex, PARP1 and translesion synthesis (TLS) DNA polymerase theta (polθ) in regulating MMEJ in higher organisms (Sfeir and Symington 2015) **(Fig. 10)**.

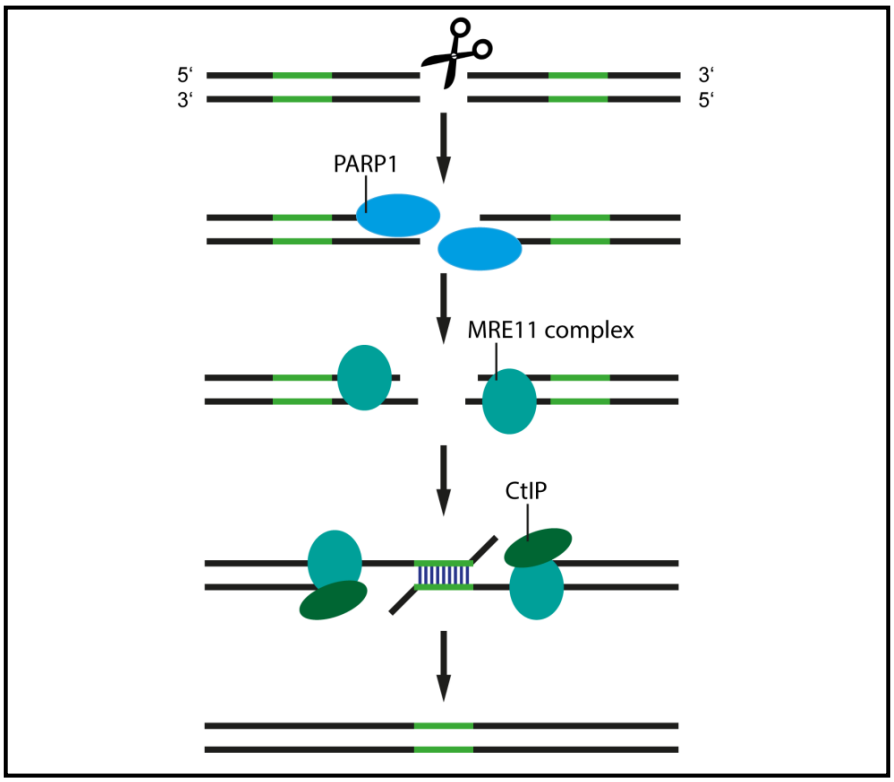

Fig. 10: Microhomology-mediated end joining DNA repair pathway. In the presence of microhomology region in the DNA, the DSBs can also be repaired via MMEJ. PARP1, MRE 11 complex, and CtIP plays an important role in MMEJ DNA repair.

1.3.5 DNA donor templates

The efficiency of HDR can be influenced by a lot of endogenous and exogenous factors, including the cell cycle, the cell type, chromosomal region, the activity of the repair system, and the DNA donor template (Yanik, Muller et al. 2016). To enhance HDR, different approaches have been developed, including the manipulation of the cell cycle and the regulation of expression of key repair pathway proteins (Chu, Weber et al. 2015, Maruyama, Dougan et al. 2015,Srivastava and Raghavan 2015). However, these invasive manipulations may be undesirable for therapeutic applications because they can alter the cellular response to DNA damage at other non-target sites in the genome and lead to tumor formation.

In contrast, designing optimal DNA donor templates can increase HDR frequencies and at the same time leave cell cycle regulation untouched. Linearized or double-stranded DNA plasmid sequences, as well as ssDNA oligonucleotides, are used as template for homologous recombination at the target site **(Fig. 11)**. Viral vectors such as AAV or IDLV can also be used as a source of donor DNA, which provide single-stranded DNA as template for HDR (Hirsch, Green et al. 2010, Handel, Gellhaus et al. 2012, Coluccio, Miselli et al. 2013, Genovese, Schiroli et al. 2014).The size of the intended sequence changes, the length of the homology arms, and the insertion site of the mutation are important factors to be considered. Although the exact mechanism by which donor design increases HDR frequencies is still under investigation, several evidences have shown its influence on gene targeting outcomes.

In mammalian cells, a plasmid donor with at least 1-2 kb of total homology is usually used for creating large sequence changes in the presence of target cleavage, including insertion of reporter genes such as fluorescent protein or antibiotic resistance markers (Dickinson, Ward et al. 2013, Yang, Wang et al. 2013). Without target cleavage, a total of 8-15 kb homology is normally used (Wu, Ying et al. 2008). Generally, the efficiency of recombination increases as the length of homology arms increases, while the efficiency decreases as the size of the DNA insert increases (Li, Wang et al. 2014), but if the homology arms contain repetitive DNA sequences, the targeting efficiency will be low (Wu, Ying et al. 2008).

In many proof-of-concept studies, it has been demonstrated that disease-causing mutation can be corrected by using DNA donor templates along with targeted nucleases. For the correction of the *IL2Rγ* gene mutation, ca. 1,543 bp centered plasmid donor has been used together with ZFN, which has shown about 7% of cells with desired genetic modification (Urnov, Miller et al. 2005).

For small sequence changes, ssDNA sequences are usually more efficient than plasmid donors. To correct the duchenne muscular dystrophy gene (*Dmd*) mutation in the germ line of *mdx* mice, a single-stranded oligodeoxynucleo-tide (ssODN) has been used as a template for HDR-mediated gene repair, which contains 90 base pairs (bp) of homology sequence flanking each side of the target site (Long, McAnally et al. 2014). Four single stranded *Crb1*^{rd8} correction ssODNs (200-

mer and 52-mer, in sense and antisense directions) with homology centered to the targeted region has been compared to stimulate HDR events for correction of the *Crb1*^{rd8} allele in C57BL/6N mice, in which 200-mer sense ssODN has shown the best result (Low, Krebs et al. 2014).

Recently, enhanced HDR rates have been reported by using optimized asymmetric ssDNA donor templates for conversion of a BFP reporter gene into a GFP reporter gene via mutation of three nucleotides within the BFP reading frame (Richardson, Ray et al. 2016). In this study, it has been observed that donor DNA complementary to the nontarget strand is more effective than donor complementary to the target strand. The optimized donor DNA is complementary to the nontarget strand by overlapping the Cas9 cut site with 36 bp in the PAM-distal side, and with 91 bp on the PAM-proximal side (Richardson, Ray et al. 2016). It was shown that optimizing a donor template at the 5' and 3' homology regions flanking the DSB site could boost the frequency of HDR in the absence of chemical and genetic intervention.

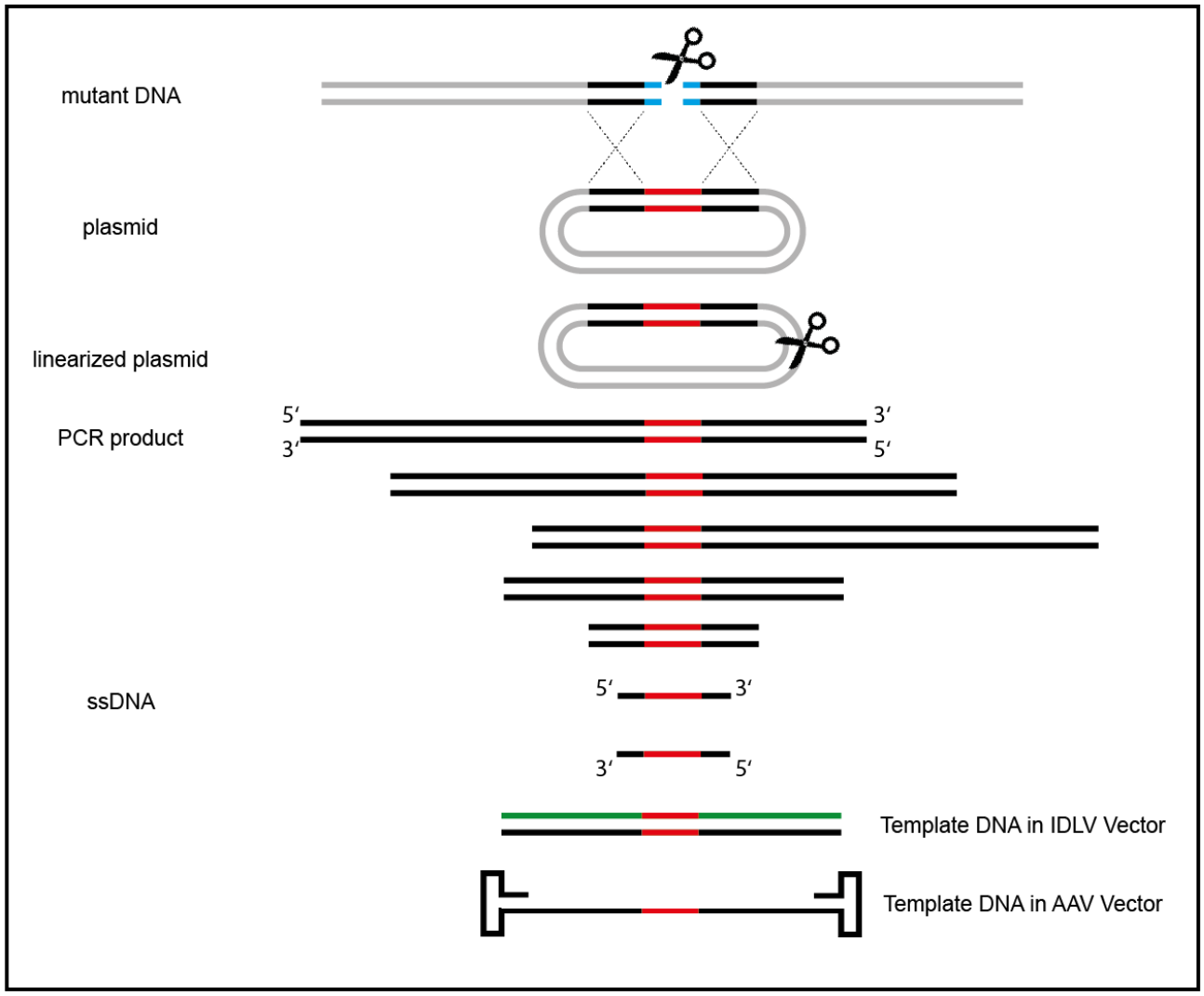

Fig. 11: Design of currently available templates. Templates can be generated as double stranded plasmid, linearized plasmid, PCR product, single stranded (ss)DNA or viral vector DNA (AAV or IDLV) (Yanik, Muller et al. 2016).

1.3.6 Methods to study DNA repair pathway choices

To quantify the DNA double strand repair outcomes, different methods have been developed including PCR amplification, followed by either direct sequencing of the modified region or, if the sequence changes containing restriction enzyme recognition sites, restriction-fragment length

polymorphism (RFLP) analysis (Ran, Hsu et al. 2013).SURVEYOR nuclease assay is also a popular method to quantify the NHEJ events, which is based on the SURVEYOR nuclease cleavage of the reannealed heteroduplexes resulted from indels of NHEJ DNA repair (Ran, Hsu et al. 2013). In addition, targeted genome modifications can be detected by deep sequencing or other next generation sequencing techniques (Ran, Hsu et al. 2013). Several reporter systems have also been developed for the direct measurement of the repair events without sequencing, including DR-GFP reporter, BFP reporter, and traffic light reporter system (Vriend, Jasin et al. 2014, Richardson, Ray et al. 2016).

In this work, the traffic light reporter (TLR) system has been used to monitor DNA repair activities, which allows rapid observation of repair pathway choices (HR or NHEJ) in cells based on fluorescence microscopy and flow-cytometric analysis (FACS).The TLR system consists of a non-functional green fluorescent protein (GFP), followed by a self-cleaving T2A peptide and a second red fluorescent protein (mCherry) in a reading frame shifted by 2 bp (Certo, Ryu et al. 2011). The GFP cDNA sequence contains an insertion comprising an I-SceI site and a stop codon, which disrupts the normal genetic function. Upon repair of the DSB induced around the stop codon, different fluorescent signals will appear depending on whether NHEJ or, in the presence of a DNA template, HDR takes place. Mutagenic NHEJ causes insertions and deletions, thus shifting the downstream mCherry sequence in frame resulting in a red fluorescent (mCherry) signal, whereas homology-directed repair restores the GFP with the help of the DNA donor template, resulting in a green fluorescent signal (Certo, Ryu et al. 2011) **(Fig. 12)**.

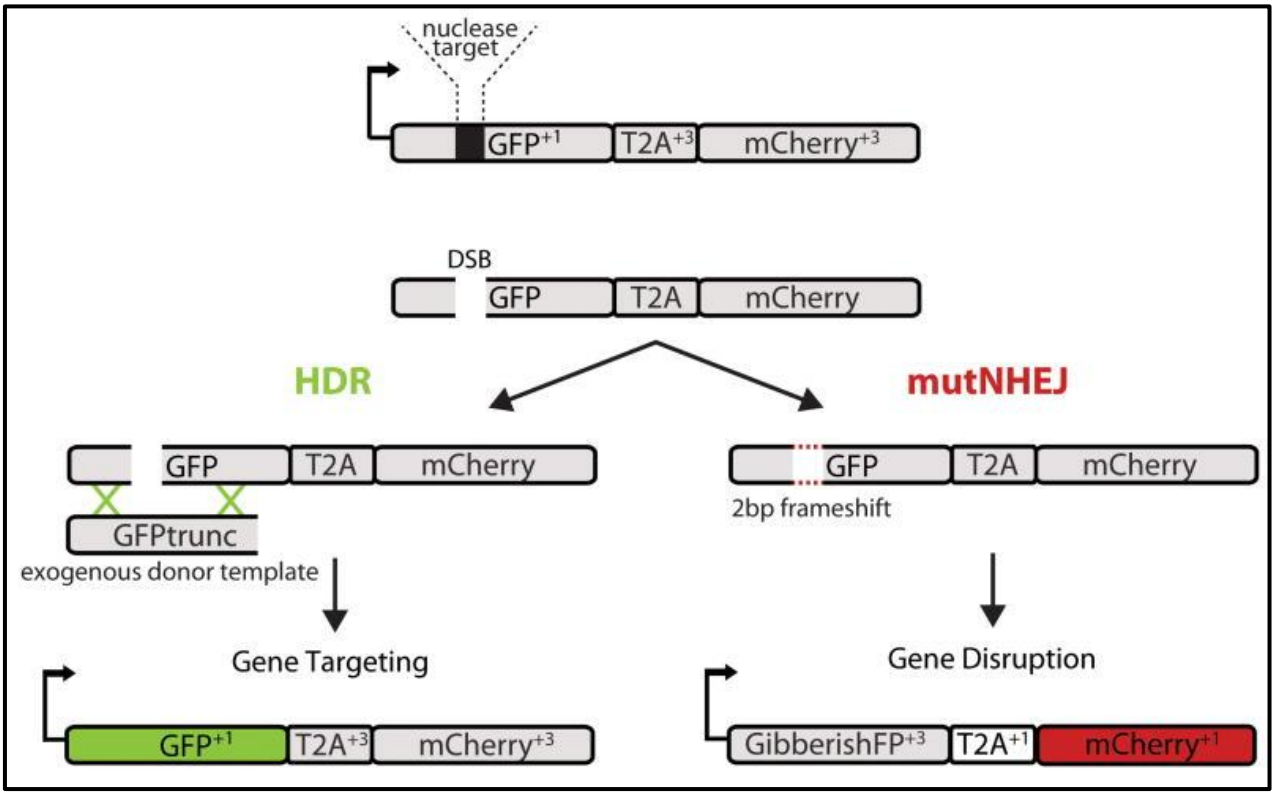

Fig. 12: The traffic light reporter system. The TLR system contains a non-functional GFP sequence, and a second mCherry sequence in a reading frame shifted by 2 bp. In the GFP cDNA sequence, an I-SceI site and a stop codon was inserted resulted in disrupting normal genetic function. Once a DSB is induced around the stop codon, it can be repaired through either NHEJ (mCherry) or in the presence of a DNA donor template, HDR (GFP) (Certo, Ryu et al. 2011).

1.4 Aim

The aim of this work is to optimize DNA double strand break repair induced by specific CRISPR-Cas9 nucleases using the Traffic Light Reporter (TLR) system for the development of homologous recombination based gene therapy. To improve the HDR rates, double-stranded DNA plasmid, linearized plasmid sequence, as well as PCR product will be used as donor template with varied homology sequence overlap on the 5’ and 3’ side of the mutation site. Furthermore, repair pathway components will be regulated through chemical or genetic manipulation to bias repair outcomes toward HDR.

2. Material and methods

2.1 Material

2.1.1 Chemicals and reagents

All chemicals and reagents listed in Table 1 were of high purity grade.

Table 1: Chemicals and reagents

Name	**Manufacturer**
Accutase	PAN
Agarose seakem LE	Biozym
Ampicillin	Sigma-Aldrich
Boric acid	Roth
Dimethylsulfoxid (DMSO)	Sigma-Aldrich
Disodium phosphate (Na_2HPO_4)	Merck
Dulbecco's Modified Eagle Medium (DMEM) high glucose	PAN
DNA loading buffer	Fermentas
Ethanol(C_2H_6O)	Roth
Ethidium bromide (EtBr)	Roth
Ethylenediaminetetraacetic acid (EDTA)	Roth
Fetal bovine serum (FBS)	PAN
GelRed	Genaxxon
Geneticin	Life Technologies
Glycerin($C_3H_8O_3$)	Merck
Hydrochloric acid (HCl)	Roth

Isopropanol (C_3H_8O)	Roth
Kanamycin	Sigma-Aldrich
L-Glutamine	PAN
Lipofectamine® LTX and PLUS™ Reagents	Invitrogen
Luria broth base	Invitrogen
Monopotassium phosphate (KH_2PO_4)	Merck
Penicillin-Streptomycin	Biochrom
Potassium chloride (KCl)	Roth
Select agar	Invitrogen
Sodium chloride (NaCl)	Merck
Sodium hydroxide (NaOH)	Roth
Tris(hydroxymethyl) aminomethane	Roth

2.1.2 Buffers

All buffers were prepared with distilled water from the Sartorius water purification system.

Table 2: Buffers

Name	**Components**
10x TBE	890 mMTris 890 mM Boric acid 20 mM EDTA Dissolved in deionized water
10x PBS	1.37 M NaCl 27 mMKCl 100 mM Na_2HPO_4 18 mM KH_2PO_4 Dissolved in deionized water, PH adjusted to 7.2 with HCl Sterilized and autoclaved

2.1.3 Media

DMEM (+++) was prepared for eukaryotic cells and LB medium was prepared for prokaryotic cells.

Table 3: Media

Name	**Components**
DMEM (+++)	DMEM 10% FBS 2 mM L-glutamine 50 U/mL Penicillin-Streptomycin
LB	25 g Luria broth base Dissolved in 1 L deionized water Autoclaved

2.1.4 Plasmids

Table 4: Plasmids and expression vectors

Name	**Provider**
pcDNA3.1 (+)	Invitrogen
pcDNA3.1 (-)	Invitrogen
pcDNA5/FRT	Invitrogen
px459 (#48139)	Addgene
hCas9 (#41815)	Addgene
gRNA Cloning Vector (#41824)	Addgene

2.1.5 Oligonucleotides

Oligonucleotides listed in Table 5 were obtained from Metabion (Planegg-steinkirchen, Germany) and used for PCR or Sanger sequencing.

Table 5: Oligonucleotides

Nr.	Name	Sequence (5'→3')
2493	TLR-2a	CGCAAATGGGCGGTAGGCGTGTACG
2494	TLR-2b	TAGAAGGCACAGTCGAGGCTGATCA
2495	TLR-S1	GAGGGTGGGCCAGGGCACGGGCAGC
2496	TLR-S2	GACCATGGGCTGGGAGGCCTCCTCC
2497	TLR-S3	GCAACTACAAGACCCGCGCCGAGGT
2498	TLR-S4	CTTGTAGGTGGTCTTGACCTCAGCG
2887	eGFP inf1a	GAGGGCGAGGGCGATGCCACCTACGGCAAGCTGAC
2888	eGFP inf1b	GCATCGCCCTCGCCCTCGCCGGACACGCTGAACTT
2907	delT inf1a	TACCCTGTTATCCCTACGCAAAAAGAGCTCACCTACGGC
2908	delT inf1b	GCCGTAGGTGAGCTCTTTTTGCGTAGGGATAACAGGGTA
2909	delTAinf1a	TACCCTGTTATCCCTACGCAAAAGAGCTCACCTACGGC
2910	delTAinf1b	GCCGTAGGTGAGCTCTTTTGCGTAGGGATAACAGGGTA
2924	dGFP 1a	ATGCAGAGGCTCCGGTGCCCGTCAG
2925	dGFP 1b	AGGGCCCGGATTCTCCTCCACGTCA
2926	dGFP inf1a	GAGAATCCGGGCCCTGTGAGCAAGGGCGAGGAGCTGTTCA
2927	dGFP inf1b	CCGGAGCCTCTGCATTTACTTGTACAGCTCGTCCATGCCG
2928	dGFP S1	AGAACACCCCCATCGGCGACGGCCC
2933	MY19FS	GAGAATCCGGGCCCTAGCGAGCTGATTAAGGAGAACATGC
2934	MY20FS	GAGCCTCTGCATTCAATTAAGCTTGTGCCCCAGTTTGCTA
2964	Cas Seq1	GTTTTAAAATGGACTATCATATGC
2965	Cas Seq2	CGCCATCCTGCTGAGCGACATCCTG
2966	Cas Seq3	GCCATTAAGAAGGGCATCCTGCAG

2967	TLR RS55f	ACTTGGCAGTACATCTACGTATTAG
2968	TLR RS55r	TTGTCGGGCAGCAGCACGGGGCCG
2969	TLR RS37f	AATGGGCGGTAGGCGTGTACGGTGG
2970	TLR RS37r	CCCTTGCTCACAGGGCCCGGATTC
2971	TLR RS73f	CGTCAATAATGACGTATGTTCCC
2972	TLR RS73r	CTTGAAGTCGATGCCCTTCAGCTCG
3464	px459 bf	GAATTCTAACTAGAGCTCGCTGATC
3465	px459 br	TGGGCCAGGATTCTCCTCGACG
3466	mRFP f	GGAGAATCCTGGCCCAGCCTCCTCCGAGGACGTCATCAAG
3467	mRFP r	CTCTAGTTAGAATTCTTAGGCGCCGGTGGAGTGGCGGCCC
3242	px459T3f	GGTGAGCTCTTATTTGCGTAGTTTTAGAGCTAGAAATAGCAA G
3243	px459T3r	TACGCAAATAAGAGCTCACCGGTGTTTCGTCCTTTCCACAAG
3244	px459T4f	GGGATAACAGGGTAATGTCGGTTTTAGAGCTAGAAATAGCA AG
3245	px459T4r	CGACATTACCCTGTTATCCCGGTGTTTCGTCCTTTCCACAAG
3246	px459T5f	GGCCACAAGTTCAGCGTGTCGTTTTAGAGCTAGAAATAGCA AG
3247	px459T5r	GACACGCTGAACTTGTGGCCGGTGTTTCGTCCTTTCCACAAG
3305	px459T6f	GCCGTCCAGCTCGACCAGGAGTTTTAGAGCTAGAAATAGCA AG
3306	px459T6r	TCCTGGTCGAGCTGGACGGCGGTGTTTCGTCCTTTCCACAAG
3307	px459T7f	GAGCGCCACCATGGTGAGCAGTTTTAGAGCTAGAAATAGCA AG
3308	px459T7r	TGCTCACCATGGTGGCGCTCGGTGTTTCGTCCTTTCCACAAG
3309	px459T8f	GGCCGGACACGCTGAACTTGGTTTTAGAGCTAGAAATAGCA AG
3310	px459T8r	CAAGTTCAGCGTGTCCGGCCGGTGTTTCGTCCTTTCCACAAG
3238	gRNAT3f	TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGG

		TGAGCTCTTATTTGCGTA
3239	gRNAT3r	GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACTAC GCAAATAAGAGCTCACC
3067	gRNAT4f	TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGG GATAACAGGGTAATGTCG
3068	gRNAT4r	GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACCG ACATTACCCTGTTATCCC
3240	gRNAT5f	TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGG CCACAAGTTCAGCGTGTC
3241	gRNAT5r	GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGA CACGCTGAACTTGTGGCC
3252	gRNAT6f	TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGC CGTCCAGCTCGACCAGGA
3253	gRNAT6r	GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACTCC TGGTCGAGCTGGACGGC
3254	gRNAT7f	TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGA GCGCCACCATGGTGAGCA
3255	gRNAT7r	GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACTGC TCACCATGGTGGCGCTC
3256	gRNAT8f	TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGG CCGGACACGCTGAACTTG
3257	gRNAT8r	GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACCA AGTTCAGCGTGTCCGGCC
3298	gesamtTLRf	GTTGACATTGATTATTGACTAGT
3299	gesamtTLRr	CAGCTGGTTCTTTCCGCCTCAGAAG
3329	RS100Af	ACTGGACGGCGACGTAAACGGCCAC
3330	RS100Ar	ACGGTGGTGCAGATGAACTTCAGGG
3417	CasSeqf	GAGGGCCTATTTCCCATGATTCC
3418	CasSeqr	GAGAGTGAAGCAGAACGTGGGGC

3419	GFPseqf	CTCCGCCCCATTGACGCAAATGGG
3420	GFPseqr	GGATGTTGCCGTCCTCCTTGAAGTC
3191	gRNAseqf	GATGCATGCTCGAGCGGCCGCCAG
3192	gRNAseqr	GAGTTAGCTCACTCATTAGGCACC

2.1.6 Enzymes

All the restriction enzymes listed in Table 6 were used with the recommended buffers according to the manufacturer's manual.

Table 6: Restriction enzymes

Name	**Manufacturer**
BamHI	New England Biolabs (NEB)
BbsI	New England Biolabs (NEB)
DraIII	New England Biolabs (NEB)
EcoRI	New England Biolabs (NEB)
EcoRV	New England Biolabs (NEB)
HindIII	New England Biolabs (NEB)
NcoI	New England Biolabs (NEB)
NdeI	New England Biolabs (NEB)
NheI	New England Biolabs (NEB)
NotI	New England Biolabs (NEB)
XhoI	New England Biolabs (NEB)

The polymerases listed in Table 7 were used for PCR reactions.

Table 7: Polymerases

Name	**Manufacturer**
PrimeSTAR® HS DNA Polymerase	Takara
Phusion® High-Fidelity DNA Polymerase	New England Biolabs (NEB)

2.1.7 Markers

Markers listed in Table 8 were used for DNA gel electrophoresis.

Table 8: Markers

Name	**Manufacturer**
GeneRuler 100bp Plus DNA Ladder	Fermentas
GeneRuler 1kb DNA Ladder	Fermentas

2.1.8 Kits

All the kits listed in Table 9 were used for DNA purification.

Table 9: DNA purification kits

Name	**Manufacturer**
NucleoSpin® Plasmid	Macherey-Nagel
QIAfilter Plasmid Midi Kit	Qiagen
Qiagen Plasmid Maxi Kit	Qiagen
NucleoSpin® Gel and PCR Clean-up	Macherey-Nagel
PureLink® Genomic DNA Mini Kit	Invitrogen

All the kits listed in Table 10 were used for DNA cloning.

Table 10: DNA cloning kits

Name	**Manufacturer**
In-Fusion® HD Cloning Kit	Clontech
TOPO® TA Cloning® Kit	Invitrogen
Zero Blunt® TOPO® PCR Cloning Kit	Invitrogen

2.1.9 Bacterial strains

Table 11: Bacterial strains

Name	**Manufacturer**
One Shot® TOP10 Chemically Competent *E. coli*	Invitrogen
One Shot® TOP10 Electrocomp™ *E. coli*	Invitrogen
Stellar™ Competent Cells	Clontech

2.1.10 Devices

Table 12: Devices

Name	**Manufacturer**
Incubator	Binder
Autoclave DX-65	Systec GmbH
Microscopy	VWR
Laminar air flow	Invitrogen
BioDocAnalyze	Biometra
BioPhotometer	Eppendorf
BD Canto II	BD
BD Aria II	BD

Centrifuge 1-15 PK	Sigma
Centrifuge AK15	Sigma
CO_2 incubator	Binder
Electrophoresis power supply	Biometra
Electrophoresis chambers	Biometra
Fluorescence Microscopy	Keyence
Gel chamber Whatman	Biometra
Thermoblock	Biometra
Vortex	VWR
Ice machine	Scotsman
Magnetic stirrer	IKA
PCR-Cycler T Professional Basic Gradient	Biometra
PH meter	Mettler-Toledo GmbH
Shaker Certomat H	Sartorius
Thermoblock	Biometra
Scale	Ohaus
Water bath TW12	Julabo

2.2 Methods

2.2.1 PCR

PCR (Polymerase chain reaction) is a biological molecular method to amplify a particular DNA sequence through the thermal cycling. The basic components of the PCR include:

1. DNA template, containing the target DNA region to be amplified;
2. two primers, complementary to the 3’ ends of the DNA target region;

3. DNA polymerase, enabling the amplification;
4. deoxynucleotide triphosphates (dNTPs), the buliding-blocks of the new synthesized DNA strand;
5. buffer solution, providing a suitable environment for the DNA polymerase and the reaction;
6. manganese ions, an essential cofactor for the DNA polymerase.

Through a series of repeated temperature changes (cycle), the target region of the DNA template can be amplified. A PCR cycle consists of a denaturation step, an annealing step and an elongation step. The denaturation step is preceded by heating the reaction to a temperature of 94-96 °C which causes DNA melting of the DNA template, yielding single-stranded DNA molecules. It is then followed by an annealing step at 50-65 °C which allows the primers annealing to the single-stranded DNA template. After the primers anneal to the DNA template, the DNA polymerase synthesizes a new DNA strand by adding dNTPs. The new DNA stand is then elongated from 5' to 3' direction commonly at a temperature of 72 °C. The reaction is finally held at 4 °C for a short-term storage. The tables listed below are approach and program for Phusion polymerase, which has proof-reading activity.

Table 13: PCR approach

PCR approach for 25 µl	
ddH2O	16.25 µl
HF Buffer (5x)	5 µl
dNTP (10 mM)	0.5 µl
Primer f (10 pmol)	0.5 µl
Primer r (10 pmol)	0.5 µl
Phusion Polymerase	0.25 µl
DNA (100 ng/µl)	2 µl
Total	25 µl

Table 14: PCR program

PCR program		
Cycles	Temperature	Time
1	98 °C	30 sec
35	98 °C	10 sec
	65 °C	30 sec
	72 °C	60 sec
1	72 °C	10 min
1	10 °C	10 min

2.2.2 DNA gel electrophoresis

To analyze the size of the DNA fragments, agarose gel electrophoresis was used. The samples were first mixed with 5x loading buffer and loaded on 1% (w/v) agarose gel in 1x TBE buffer containing 0.00001% GelRed® (Genaxxon). The run was performed in 1x TBE buffer and the bands were visualized and documented using Biodoc Analyze (Biometra).

2.2.3 Restriction enzyme digestion

Restriction enzyme digestion has been used to cut the DNA at specific sites. The digestion was performed using restriction enzymes purchased from New England Biolabs, according to the manufacture's instruction. The DNA fragments were visualized using gel electrophoresis.

2.2.4 Gel extraction

The PCR products were sliced from the agarose gel under UV light with short exposition time after gel electrophoresis. They were purified with PCR clean-up Gel extraction kit (Macherey-Nagel) and finally eluted with ddH_2O. The ddH_2O was then used as blank value for the measurement of the concentration.

2.2.5 Plasmid DNA isolation

Low-scale plasmid DNA was isolated using NucleoSpin® Plasmid QuickPure (Macherey-Nagel) kit and high-scale plasmid DNA was isolated using Qiagen® Plasmid Midi or Maxi (Qiagen) kit, according to the manufacturer's instruction.

2.2.6 Genomic DNA isolation

The cells were first harvested in a 50 ml falcon by centrifugation at 1150 rpm (revolutions per minute) for 10 min at 4 °C. The supernatant was discarded and the cell pellet was resuspended in 10 ml 0.15 M KCl. The solution was then centrifuged at 1150 rpm for 10 min at 4 °C. After this the supernatant was discarded and the pellet was dissolved in 5 ml SE buffer, 25 µl Proteinase K (10 mg/ml) and 250 µl 20% SDS at 55 °C for 3 h. 1.4 ml 6 M NaCl was then added to the cell lysis, the solution was vortexed and centrifuged at 5100 rpm for 20 min at 20 °C. The supernatant containing the DNA was transferred to a new falcon and 2 Vol ice cold absolute EtOH was added. The solution was then inverted and centrifuged at 5100 rpm for 30 min at 4 °C. The DNA pellet was then washed with 70% EtOH and dissolved in 100 µl ddH_2O. The ddH_2O was used as blank value for the measurement of the DNA concentration.

2.2.7 Transformation

2.2.7.1 Electroporation

Electro-competent cells were first thawed on ice and then gently tapped to ensure that the cells were suspended. 2.5 µl plasmid DNA or cloning reaction mixture were added to the cells and gently mixed for the evenly distribution. The tube was leaved on ice for 5-30 min. The mixture was then transferred into a cold electroporation cuvette. The transformation was performed in an electroporator at 1250 V, 25 mA, and 25 Ω. The cells were resuspended in 1 ml SOC medium and transferred into a 1.5 ml tube. The cells were incubated at 37 °C for 60 min while shaking at 250 rpm. 50 µl of the cells were spread on LB plate containing 100 µg/ml of ampicillin or other antibiotic for the cloning vector.

2.2.7.2 Chemical transformation (heat shock)

Chemical-competent cells were first thawed on ice and then gently tapped to ensure that the cells were suspended. 2.5 µl plasmid DNA or cloning reaction mixture were added to the cells and gently mixed for the evenly distribution. The tube was leaved on ice for 30 min. The cells were heat shocked in a heat block at 42 °C for 45 sec, and then placed directly on ice for 1 min. After a heat shock, 450 µl SOC medium was added to the cells. The cells were incubated at 37 °C for 60 min while shaking at 250 rpm. 50 µl of the cells were spread on LB plate containing 100 µg/ml of ampicillin or other antibiotic for the cloning vector.

2.2.8 Generation of CRISPR-Cas9 Targets and Cas9-mRFP constructs

The Cas9 expression vector, pSpCas9(BB)-2A-Puro (PX459) V2.0, was purchased from Addgene (Cambridge, MA, plasmid 62988). The Cas9 target was cloned into the px459

expression vector using In-Fusion Cloning Kit (Clontech, USA), according to the manufacturer's instructions. The puromycin resistance gene was exchanged with mRFP gene using In-Fusion Cloning Kit. The plasmid containing mRFR gene was purchased from Addgene (Cambridge, MA, plasmid 13032). Plasmids were prepared by using the Qiagen Maxi plasmid kit (Qiagen, Germany).

2.2.9 Design and generation of donor templates

For the generation of the donor templates, the mutant GFP gene of the TLR3 system was corrected with wide type GFP gene sequence by using In-Fusion Cloning Kit (Clontech, USA), according to the manufacturer's instructions. PCR donor templates were amplified using Phusion polymerase (NEB, Germany) and the PCR products were further cloned into the TOPO TA Cloning Kit (Life Technologies, USA) for the generation of the plasmid donor templates. Linearized plasmid donor templates were generated by digesting the plasmids using HindIII and DraIII endonuclease (NEB, Germany). The PCR products and linearized plasmid DNA were purified using NucleoSpin® Gel and PCR Clean-up Kit (Macherey-Nagel, Germany). Plasmids were prepared by using the Qiagen Maxi plasmid kit (Qiagen, Germany).

2.2.10 Cell culture and transfection

A stable cell line of HEK293 expressing TLR3 was generated following transfection of a pcDNA3.1(-)-TLR3 vector (Life Technologies, Catalog nos. V795-20) and selected using neomycin (Life Technologies, Geneticin®, USA). The cell line was maintained in DMEM supplemented with 10% FBS, 2 mM L-glutamine, 50 U/mL Penicillin-Streptomycin (Life Technologies, USA), and 500 µg/ml Geneticin® and cultured at 37°C with 5% CO2 incubation. Transfection of CRISPR-Cas constructs and donor templates was performed using Lipofectamine LTX (Life Technologies, USA), with 500 ng plasmid and 500 ng donor templates per well of 24-

well plate, according to the manufacturer's instructions. Fluorescence images were visualized 3 days post-transfection with a Keyence microscope (Keyence, Germany) to control the transfection efficacy.

2.2.11 Fluorescense-activated cell sorting (FACS) analysis

Cells were collected 3 days after transfection by trypsinization. The supernatant medium was first collected in 1.5 ml tube. 200 µl Accutase (PAN-Biotech, Germany)was then added to the cells and the cells were incubated at 37°C for 5-10 min. 300 µl DMEM+++ was added to the cells to stop the typsinization reaction and the cells was transferred to the tube containing supernatant medium. The cells were centrifugated at 300 g for 3 min at 4°C and the cell pellets were washed one time with 300 µl PBS. The cells were finally resuspended in 200 µl PBS for FACS analysis. FACS analysis of mRFP/BFP/GFP-positive events was performed using BD Canton II flow cyctometer (BD Biosciences, Germany). Forward (FSC) versus side (SSC) scatter plots are used to select the cells. Typically, 50,000 cells per sample are analyzed.

3. Results

3.1 Project strategy

In this study, the traffic light reporter (TLR) system will be used to monitor DNA repair activities, which allows rapid observation of repair pathway choices in cells based on fluorescence microscopy and flow-cytometric analysis (FACS). The modified TLR system comprises a bicistronic expression system of a GFP and a second fluorescent protein BFP. The GFP cDNA sequence contains an insertion of a stop codon. Upon repair of the DSB induced close to the stop codon, different fluorescent signals will appear depending on whether HDR (green signal) or NHEJ (blue signal) takes place.

The site-specific DSBs will be induced by employing the RNA-guided CRISPR-Cas9 system. To understand the influence of DNA donor templates on the results of repair pathway choices, double-stranded DNA plasmids, linearized plasmids, as well as PCR products will be used and the length of the 5' and 3' homology arms will be altered. Furthermore, plasmid donors are linearized at different sites in the backbone for the generation of differing 5' and 3' overhangs to investigate whether this also influences HDR activity. To improve the HDR rates, repair pathway components will be regulated through chemical or genetic manipulation to bias repair outcomes toward HDR **(Fig. 13)**.

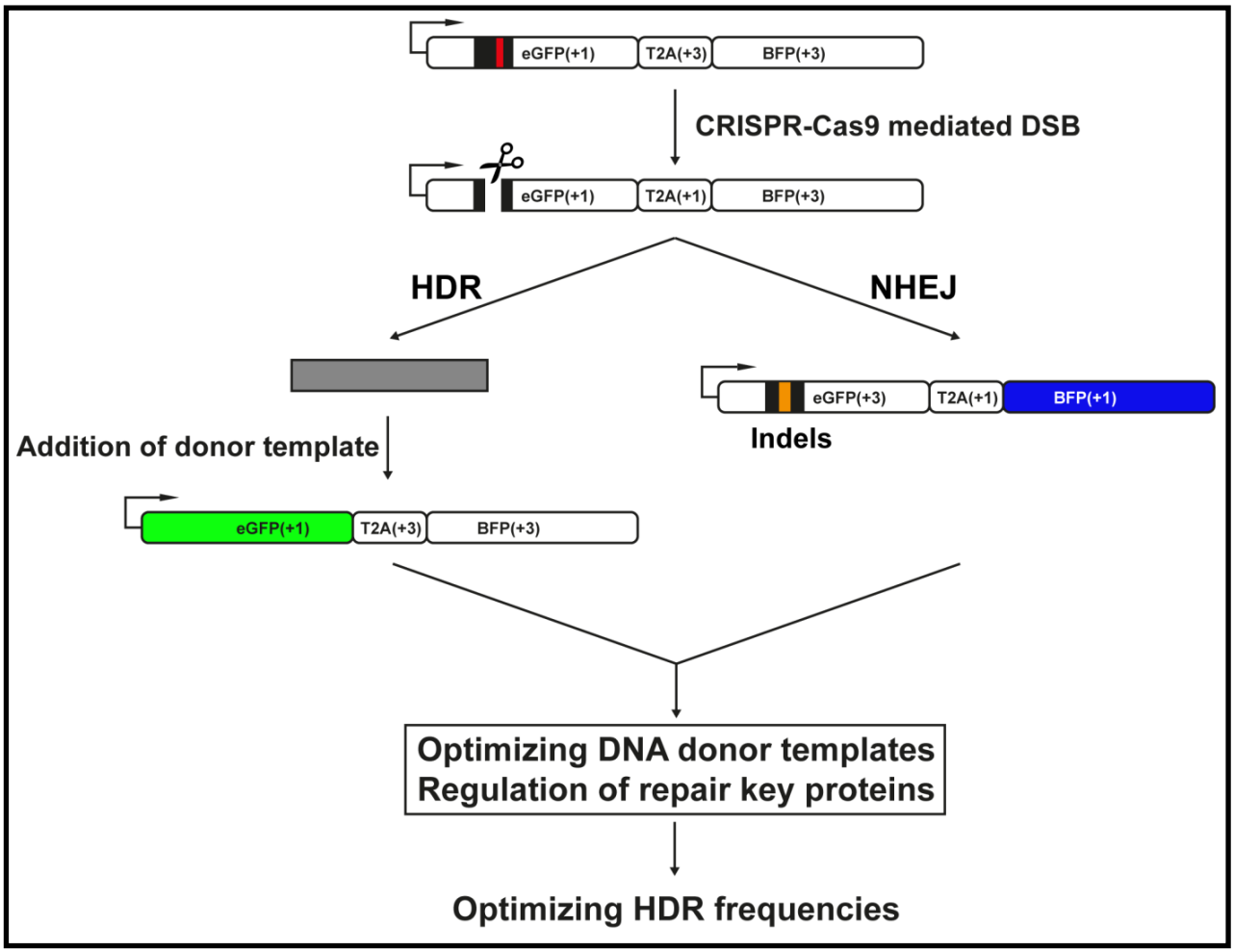

Fig. 13: Optimizing DNA double strand break repair in modified TLR3 system. Different CRISPR-Cas9 targets around the stop codon should be designed for the specific cleavage within the GFP sequence in the TLR3 system. Depending on the addition of the donor templates, DSBs can be repaired through either NHEJ (BFP) or HDR (GFP). To improve the HDR efficiency, different kinds of donor templates should be tested and repair key proteins should be regulated. Red line: stop codon. Yellow line: Indels. Black rectangle: mutation site. Scissors: site specific CRISPR-Cas9 nucleases.

3.2 Establishment of the TLR systems

3.2.1 Generation of the different TLR systems

To enable the observation and analysis under different lasers from fluorescence microscopy and flow cytometry, the second fluorescent protein (mCherry) of the TLR1 system was exchanged with green fluorescent protein (GFP) (TLR2) and blue fluorescent protein (BFP) (TLR 3) **(Fig. 14)**.

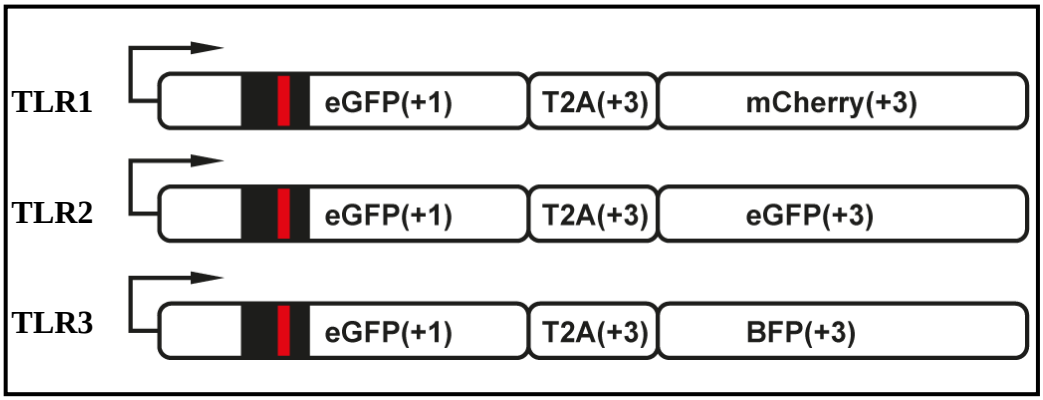

Fig. 14: TLR1, TLR2, and TLR3 system. The original TLR system (TLR1) consists of a mutated GFP sequence, T2A peptide and a second red fluorescent protein (mCherry). The second fluorescent gene of modified TLR systems, TLR2 and TLR3, was exchanged with GFP and BFP. Red line: stop codon. Black rectangle: mutation site.

The original TLR plasmid (TLR1) was purchased from Addgene (Cambridge, MA, plasmid #31482) with lentiviral vector backbone. For the in vitro usage in HEK293 cells, the TLR1 cDNA sequence was cloned into the pcDNA5/FRT, pcDNA3.1(+), or pcDNA3.1(-) vector (Thermo Fisher Scientific, USA) under the CMV promoter for high level expression of the gene of interest. For the generation of the TLR2 and TLR3 systems, functional GFP and BFP sequences were first amplified without ATG start codon and then cloned into the backbone PCR

product of TLR1 without mCherry sequence using the In-Fusion Cloning Kit (Clontech, USA) **(Fig. 15)**. The plasmids were examined with restriction digestion and Sanger sequencing (Data not shown). In this work, TLR3 was used in the cell culture to optimize double strand repair pathway choices.

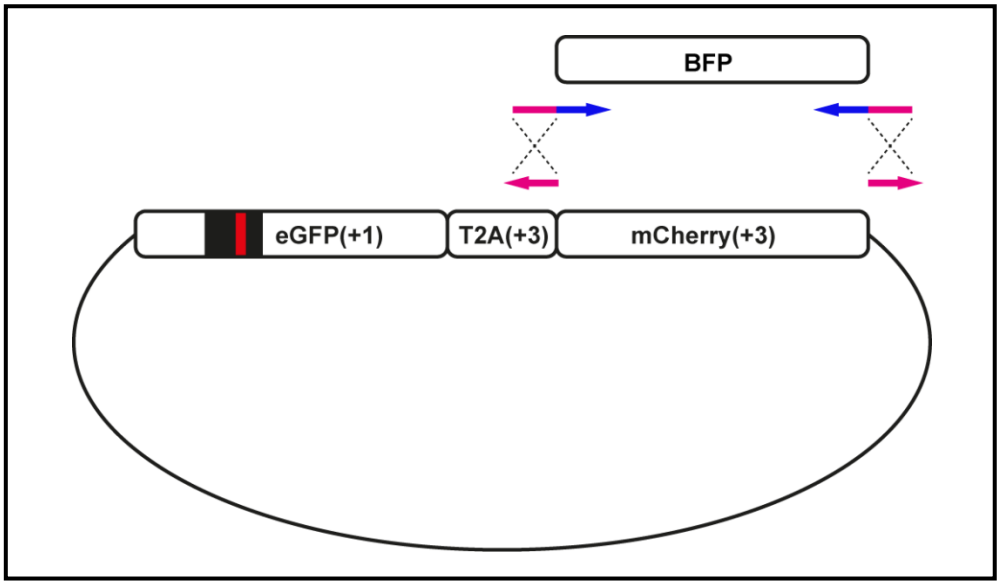

Fig. 15: Cloning strategy of TLR3. pcDNA5/FRT-TLR1 plasmid was first amplified without mCherry sequence (PCR primer in pink with arrow). BFP cDNA sequence was then amplified without ATG start codon, but with 15 bp extensions homologous (PCR primer in blue with arrow and pink overhang) to vector ends. Upon using the In-Fusion Cloning Kit, the two PCR products are fused due to the 15 bp homology for the generation of TLR3 plasmid. Red line: stop codon. Black rectangle: mutation site.

3.2.2 Generation of NHEJ and HDR controls for the TLR systems

To establish the TLR systems, artificial NHEJ control (TLR-delTA) was first created through deletion of 2 nucleotides (thymin and adenin) of the stop codon within the mutated GFP sequence, which removes the stop codon and shifts the second fluorescent gene in frame. In order to create the artificial HDR control, the mutated GFP sequence was corrected with the help of wild type

GFP sequence for the generation of the functional GFP gene. All the TLR systems, NHEJ controls and HDR controls are under the CMV promoter and BGH polyA signal with pcDNA5/FRT backbone **(Fig. 16)**. The plasmids were examined through restriction digestion and Sanger sequencing **(Fig. 17)**.

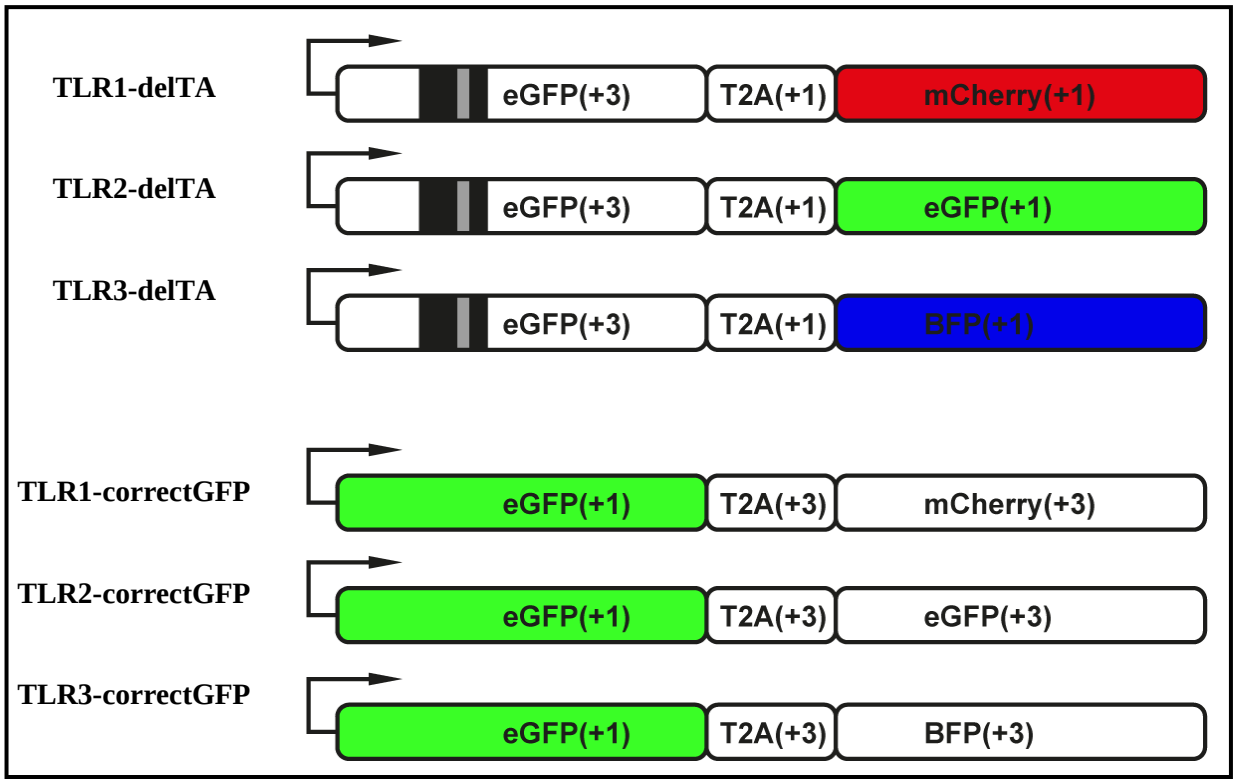

Fig. 16: NHEJ and HDR controls of TLR1, TLR2, and TLR3 system. NHEJ and HDR control for the TLR systems were generated through deletion of 2 bp nucleotides (thymine and adenine) and correction of mutated GFP sequence, respectively. Grey line: 2 bp deletion.

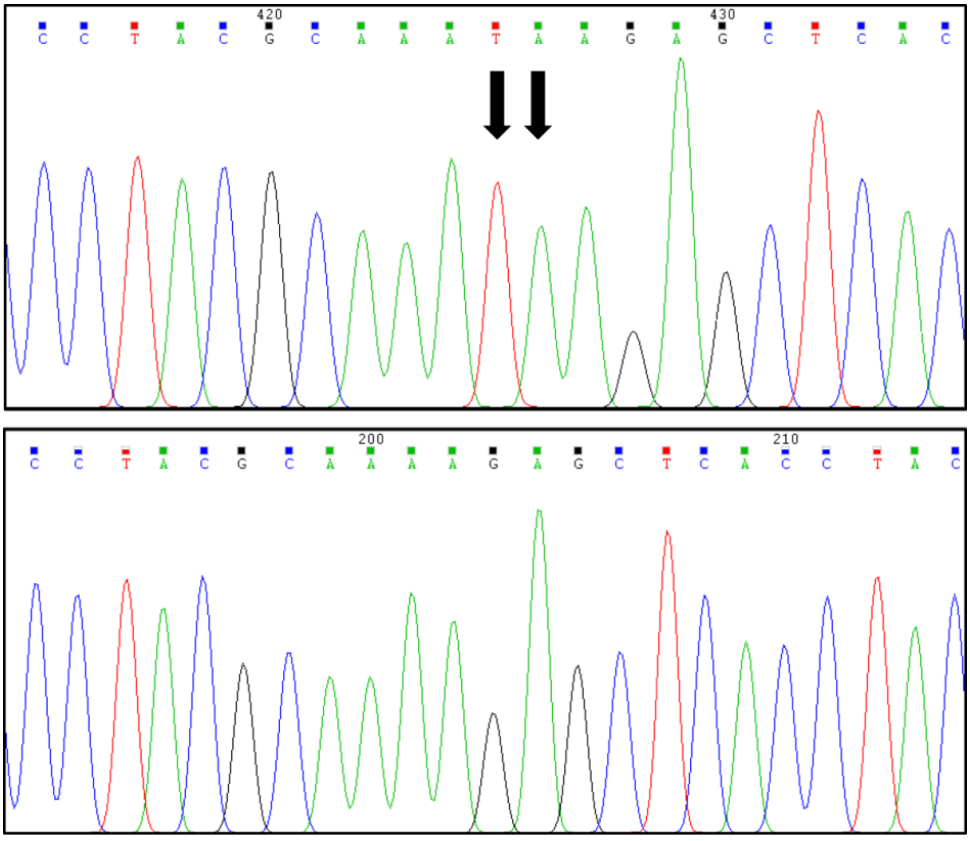

Fig. 17: Comparison of TLR and TLR-delTA sequence. 2 bp of stop codon within the mutated GFP sequence were deleted for the generation of the NHEJ control of the TLR systems. Black arrow shows the deleted thymine and adenine of the stop codon (TAA). The top image shows the Sanger sequencing result of the TLR3 system around the stop codon and the bottom image shows the TLR3-delTA control.

After the TLR systems and control plasmids were cloned into the pcDNA5/FRT vector, the expression of the TLR systems was examined. The transfection efficiency was first optimized with different amounts of DNA, different amounts of transfection reagents, and different incubation times after the transfection (data not shown).

TLR3 plasmid was transfected into the HEK cells, and observed 72 h after transfection with the fluorescence microscope. The transfected cells did not show any fluorescence signal under the

BFP and GFP channel, due to the fact that neither the GFP gene nor the BFP gene is expressed because of the presence of the stop codon within the mutated GFP sequence. The cells were than harvested and analyzed by FACS under BFP and GFP scatter, which shows no detectable GFP signal but few BFP positive cells. The appearance of the BFP positive signal is probably because of the presence of the intact BFP cDNA sequence, which leads to trans-activation of the second BFP gene of the TLR3 system.

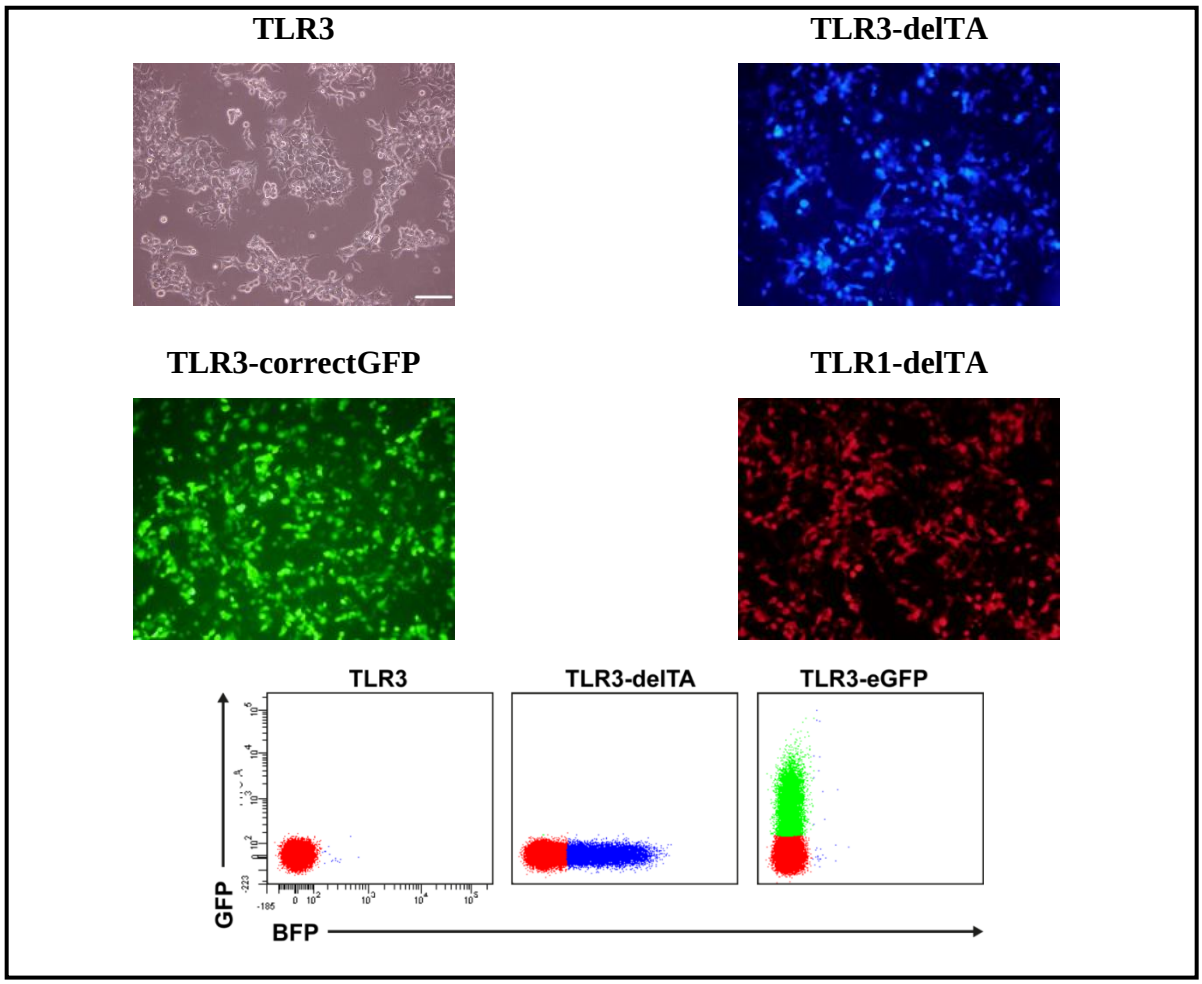

Fig. 18: Microscopy and FACS analysis of TLR control plasmids. TLR3, TLR3-delTA, TLR3-eGFP, and TLR1-delTA was transfected into the HEK 293 cells. 72 h after the transfection, the cells were observed under the microscopy and analyzed using flow cyctometry. Scale bar: 100 µm.

Besides, TLR3-delTA and TLR3-correctGFP plasmids were transfected for the examination of the NHEJ and HDR control of the system. Through 2 bp deletion of the TLR3 system, the TLR3-delTA variant shows the blue fluorescent signal and TLR1-delTA shows red fluorescence signal (mCherry). The correction of the mutated GFP sequence leads to the expression of the GFP gene, which can be observed by microscope and measured by FACS **(Fig. 18)**.

3.3 Design of different CRISPR-Cas9 targets

3.3.1 Location of six CRISPR-Cas9 targets

To induce the site specific cleavage in the TLR3 system for the stimulation of the DNA repair mechanisms, six different target sites were identified upstream or downstream of the stop codon within the GFP sequence containing the initiating 5'G and the 3'PAM (NGG), among which T3 and T4 are inside the mutated GFP sequence. Furthermore, T5 and T7 are located on the transcriptionally active strand of the TLR3 cDNA sequence, while T3, T4, T6 and T8 are located on the transcriptionally inactive strand of the cDNA sequence **(Fig. 19 and Table 15)**.

Table 15: gRNA sequences targeted TLR3 system

T3	GGTGAGCTCTTATTTGCGTAGGG
T4	GGGATAACAGGGTAATGTCGAGG
T5	GGCCACAAGTTCAGCGTGTCCGG
T6	GCCGTCCAGCTCGACCAGGATGG
T7	GAGCGCCACCATGGTGAGCAAGG
T8	GGCCGGACACGCTGAACTTGTGG

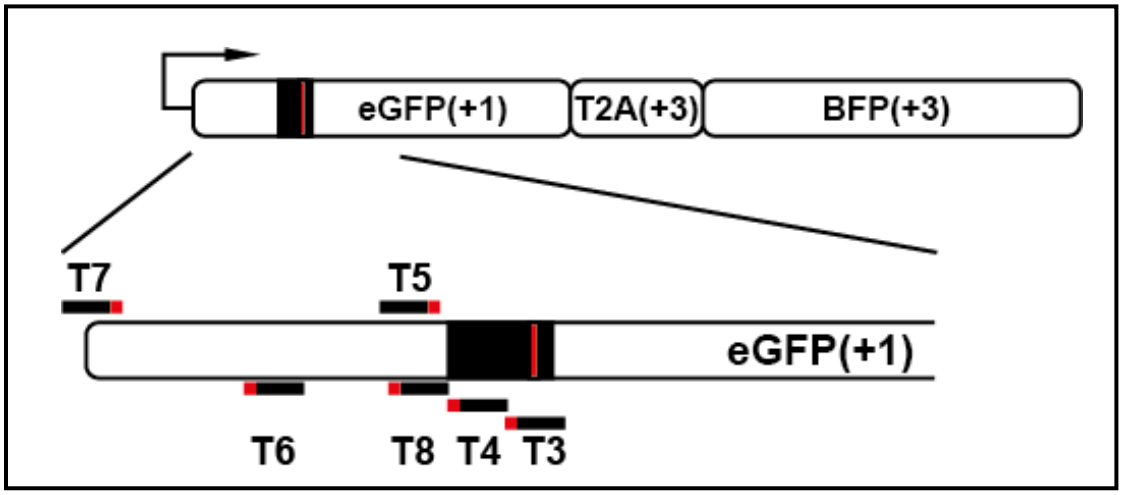

Fig. 19: CRISPR-Cas9 target sites. CRISPR-Cas9 target cleavage sites (T3, T4, T5, T6, T7, T8) are around the stop codon in the TLR sequence. 20-nt guide RNA (blue) was designed upstream of PAM sequence (NGG; red). Cas9 mediates a DSB 3 bp upstream of the PAM sequence. PAM, Proto-Spacer-Adjacent Motif; T, target; Red line: stop codon.

3.3.2 Generation of two CRISPR-Cas9 expression systems

In order to study the effect of different gRNA concentrations on cutting efficiency, two CRISPR-Cas systems were generated, one "all-in-one" vector and one system comprising two plasmids, one containing the Cas protein, and one containing the gRNA sequence **(Fig. 20)**. The respective guide RNA sequences were cloned into the px459 (Addgene, Cambridge, MA, plasmid #48139) expression vector and gRNA cloning vector (Addgene, Cambridge, MA, plasmid #41824). The px459 produces the guide RNA and the Cas9 protein simultaneously, while the gRNA cloning vector must be coexpressed with an additional hCas9 (Addgene, Cambridge, MA, plasmid #41815) expression vector for effective cleavage. Both of the two systems are able to induce DNA double strand breaks.

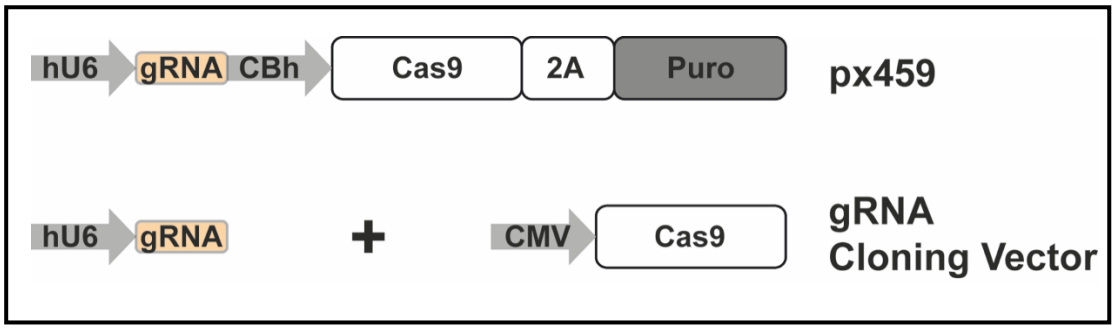

Fig. 20: Two CRISPR-Cas9 expression systems. 20 nt guide RNA sequences of the six targets were cloned into "All-in-one" vector (px459) and gRNA cloning vector to produce the guide RNA and Cas9 protein either simultaneously or separately.

3.4 Activity Test of CRISPR-Cas9 targets

After the gRNAs were cloned into the two CRISPR-Cas9 expression systems, the activity of the nucleases was tested through cotransfection with the TLR3 plasmid in HEK293 cells. After 72 hours, the transfected cells were observed under fluorescence microscopy and prepared for FACS analysis. By measuring the BFP positive cells, all six targets in both systems showed BFP positive signals compared with T0, which indicates that NHEJ takes place **(Fig. 21)**.

While the absolute values varied depending on the transfection efficiency in each experiment, the overall outcome that T5 and T7 resulted in highest levels of NHEJ was similar in both CRISPR-Cas9 expression systems. Interestingly, the latter 4 target sites with lower NHEJ activity are all located on the transcriptionally inactive minus strand of the DNA, while the two sites with higher activity are located on the transcriptionally active plus strand of the DNA, indicating that NHEJ rates are dependent on the location of the target site being on the transcriptionally active or inactive DNA strand.

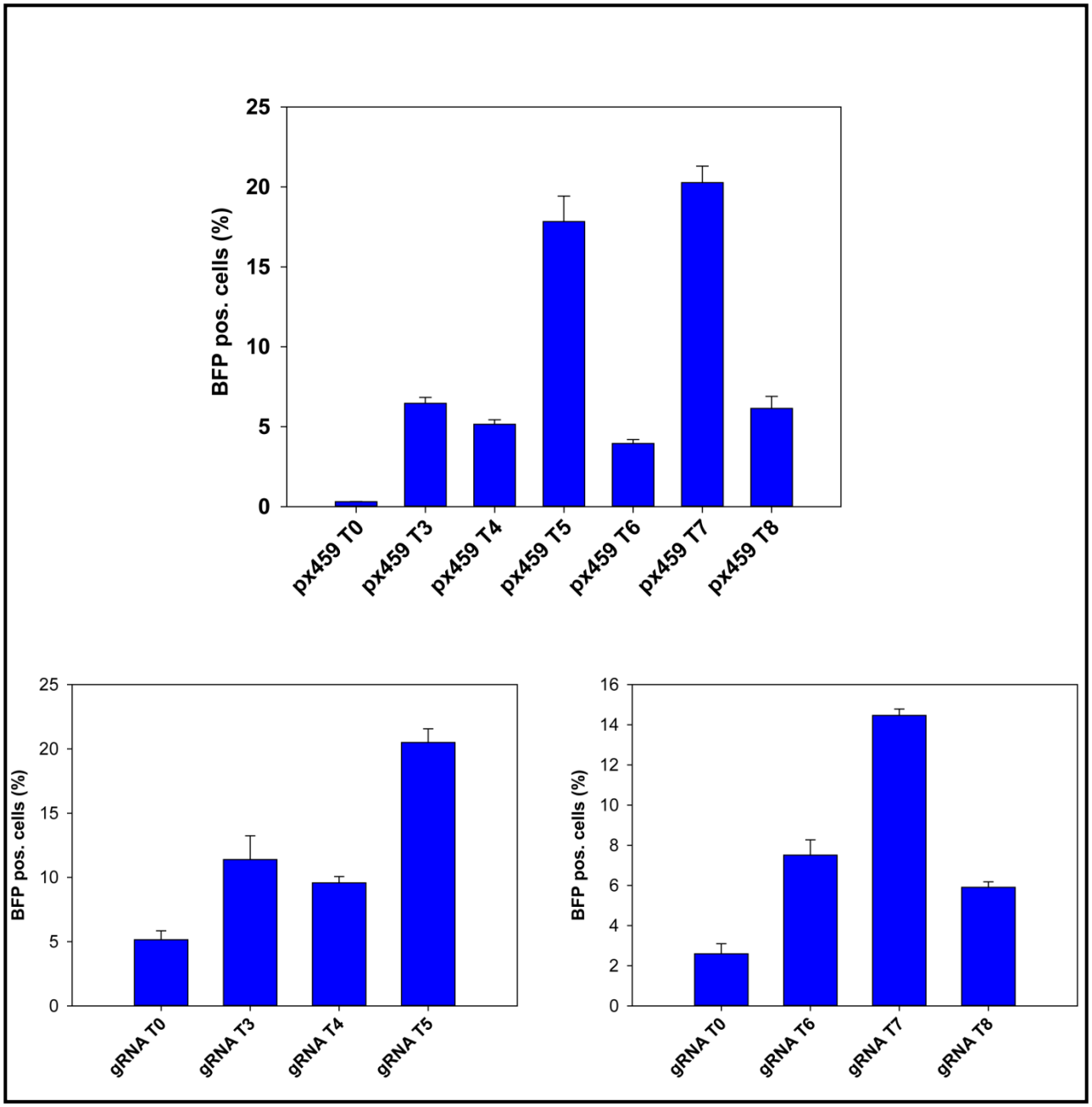

Fig. 21: Activity test of both expression systems. CRISPR-Cas9 targets were cotransfected with TLR3 plasmid in the HEK cells. The BFP positive cells were measured using flow cytometry. px459 T0: empty px459 vector. gRNA T0: empty gRNA cloning vector.

3.5 Comparison of two CRISPR-Cas9 systems

Upon cotransfection of 800 ng nucleases and 200 ng TLR3 plasmids, the efficacy of the two CRISPR-Cas9 expression systems has been compared. With the constant total amount of the nucleases, the amount of hCas9 was varied from 0 ng to 800 ng with an increase of 100 ng, while the gRNA T5decreased from 800 ng to 0 ng. To compare both Cas9 expression systems, 800 ng px459 T5 was cotransfected with 200 ng TLR3 plasmid. 72 hours after transfection, the BFP positive cells were first observed under the fluorescence microscope and counted using flow cytometry. When looking at the coexpression of hCas9 and gRNA T5, the highest activity has been shown by the cotransfection of 200 ng hCas9 and 600 ng gRNA T5, which means 1/3 of hCas9 compared to gRNA T5. With the increase of hCas9 amount, the NHEJ activity first increased until 200 ng and then decreased dramatically. Most importantly, px459 T5 has shown higher activity than coexpression of hCas9 and gRNA T5 at any concentration.

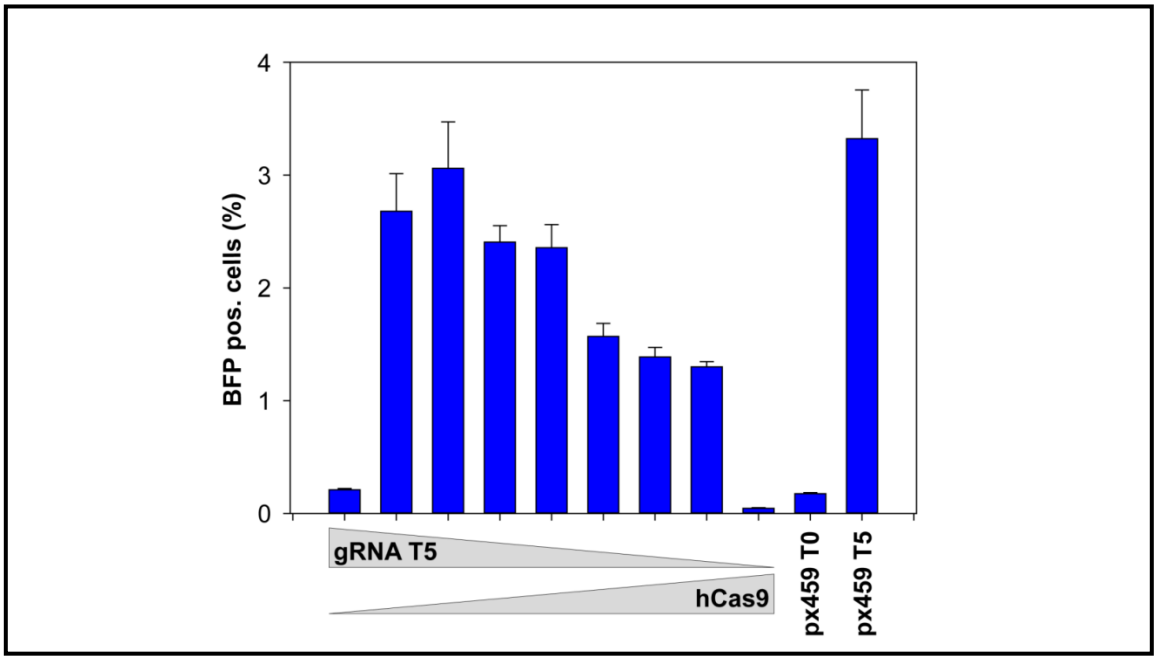

Fig. 22: Comparison of two CRISPR-Cas9 systems. Upon cotransfection of 200 ng TLR3 plasmid and 800 ng total amounts of nucleases, NHEJ activity was compared. px459 all-in-one vector has shown higher activity compared to coexpression of hCas9 and gRNA cloning vector.

3.6 Generation of the different donor templates

In order to stimulate HDR events using different kinds of DNA donor templates, 1000 bp DNA sequences were generated to correct either centrally (RS55), with a shorter 5’ homologous region (RS37), or a shorter 3’ homologous region (RS73). The different lengths of the homologous regions are 500 and 500 bp, 300 and 700 bp and vice versa. Four different kinds of DNA were employed in the experiments, uncut plasmid, linearized plasmid with 5’ or 3’ backbone overhang, or PCR product **(Fig. 23)**. TOPO TA vector backbone has been used for the generation of plasmid donor templates, which does not contain eukaryotic promoter. The templates were examined using gel electrophoresis **(Fig. 24)** and Sanger sequencing (Data not shown).

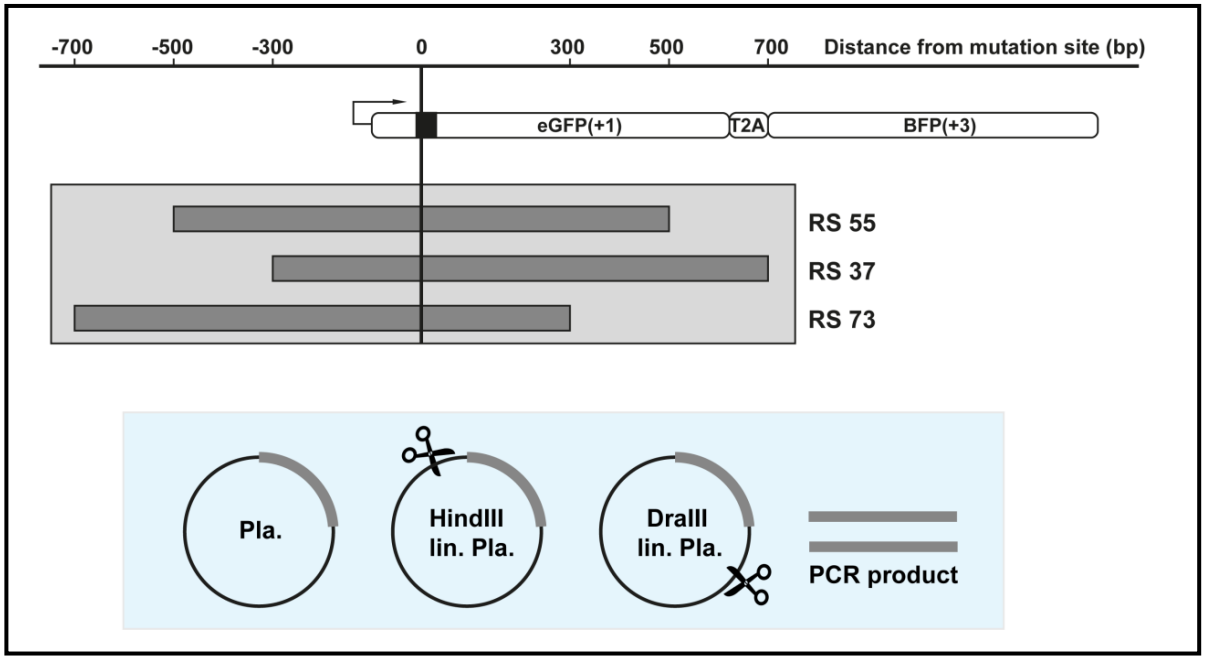

Fig. 23: Generation of different donor templates. a. 1 kb dsDNA donor templates were generated with varied homology sequence overlap on the 5’ and 3’ side of the mutation site. **b.** Donor templates were generated as plasmid, linearized plasmid with 5’ or 3’ backbone overhang, or PCR product.

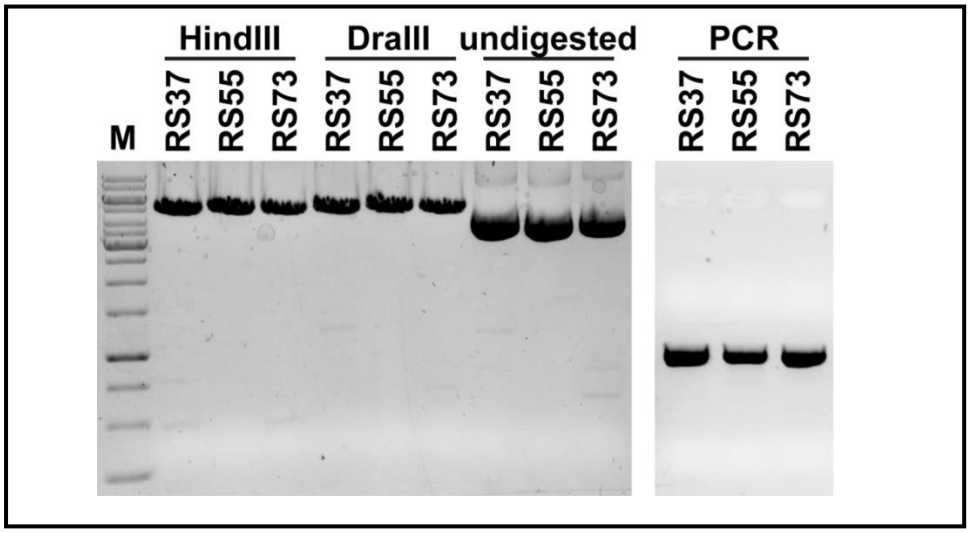

Fig. 24: Gel analysis of donor templates. Plasmid DNA containing 1000 bp RS37, RS55 and RS73 was linearized with HindIII and DraIII for the generation of the 5' and 3' backbone overhang. PCR donor templates were amplified using undigested Plasmid and Phusion polymerase.

3.7 HDR and NHEJ events using uncut plasmid RS55

In order to test HDR frequency in the TLR3 system, the HEK 293 cell line was cotransfected with the TLR3, px459 plasmid and the RS55 template plasmid, the latter containing 1000 bp of the natural GFP sequence without start region and the sequence to be repaired (i.e. the original GFP sequence that had been replaced by the I-SceI site together with the stop codon) located centrally. The application of px459 T3 and T4 resulted in a measurable HDR signal, while the application of RNA guide T5 did not show any GFP positive signal, indicating the presence of DSBs at the target site **(Fig. 25)**. This is due to the fact, that the guide RNA targets a site within the natural GFP sequence, representing a reliable control that GFP signaling in our experiments is indeed originating from the HDR activity at the target site.

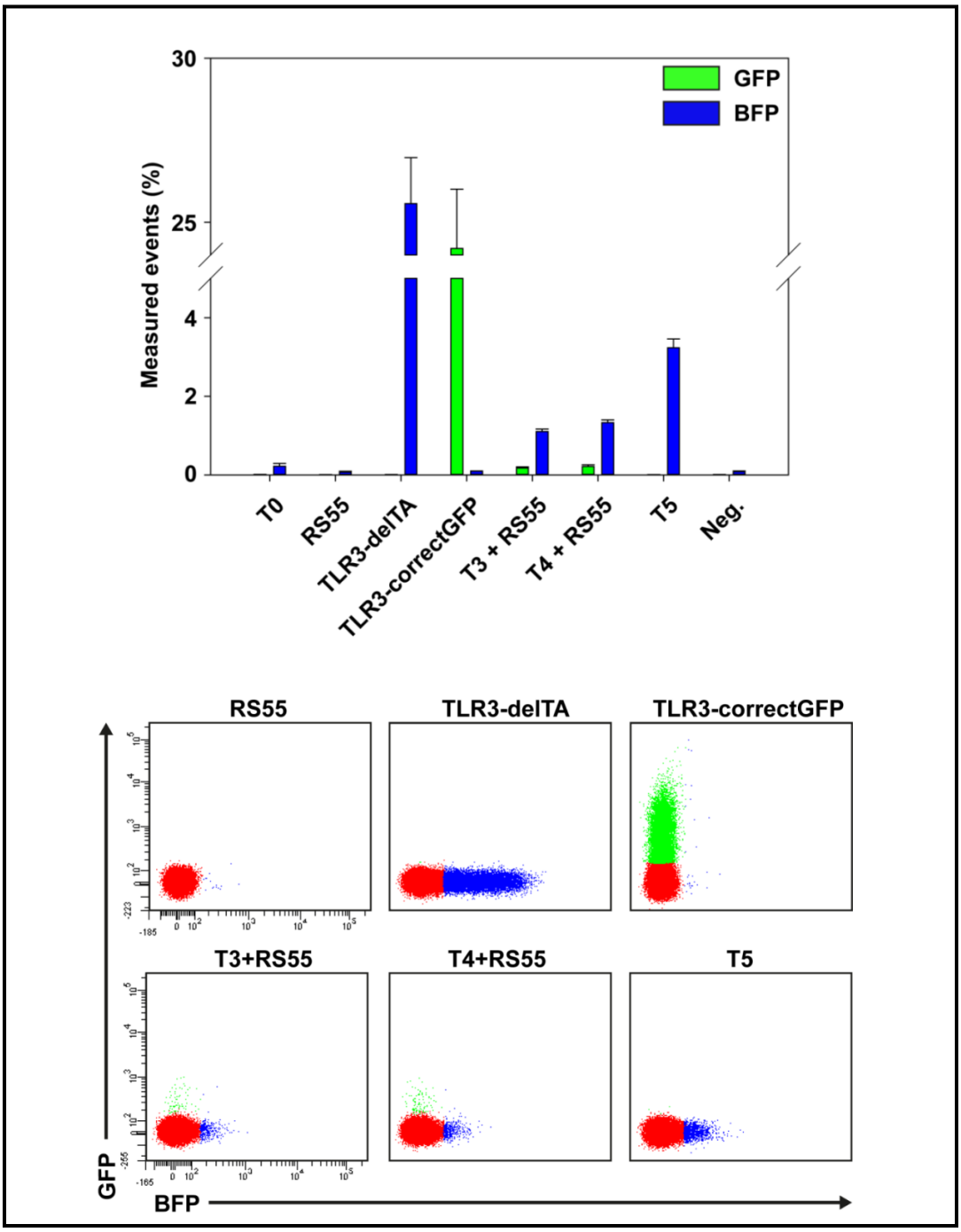

Fig. 25: FACS analysis of HDR and NHEJ events using uncut plasmid RS55. TLR3 plasmid was cotransfected with nucleases and donor template for the quantification of both HDR and NHEJ events. Artificial NHEJ (TLR3-delTA) and HDR (TLR3-correctGFP) control for the TLR systems were generated through deletion of 2 bp nucleotides (thymin and adenin) and correction of mutant sequence, respectively.

3.8 Creation of TLR3 stable cell line

To quantify the DNA repair outcomes at the genomic level, a stable HEK293-TLR3 cell line was created using the linearized pcDNA3.1(-)-TLR3 expression vector, in which the TLR3 system is under the control of a CMV promoter. The pcDNA5/FRT-TLR3 was first digested with NheI and KpnI and the TLR3 expression cassette was cloned into the pcDNA3.1(-) expression vector. The pcDNA3.1(-)-TLR3 was then linearized using PvuI, purified, and transfected into the HEK 293 cell line (Fig. 26).

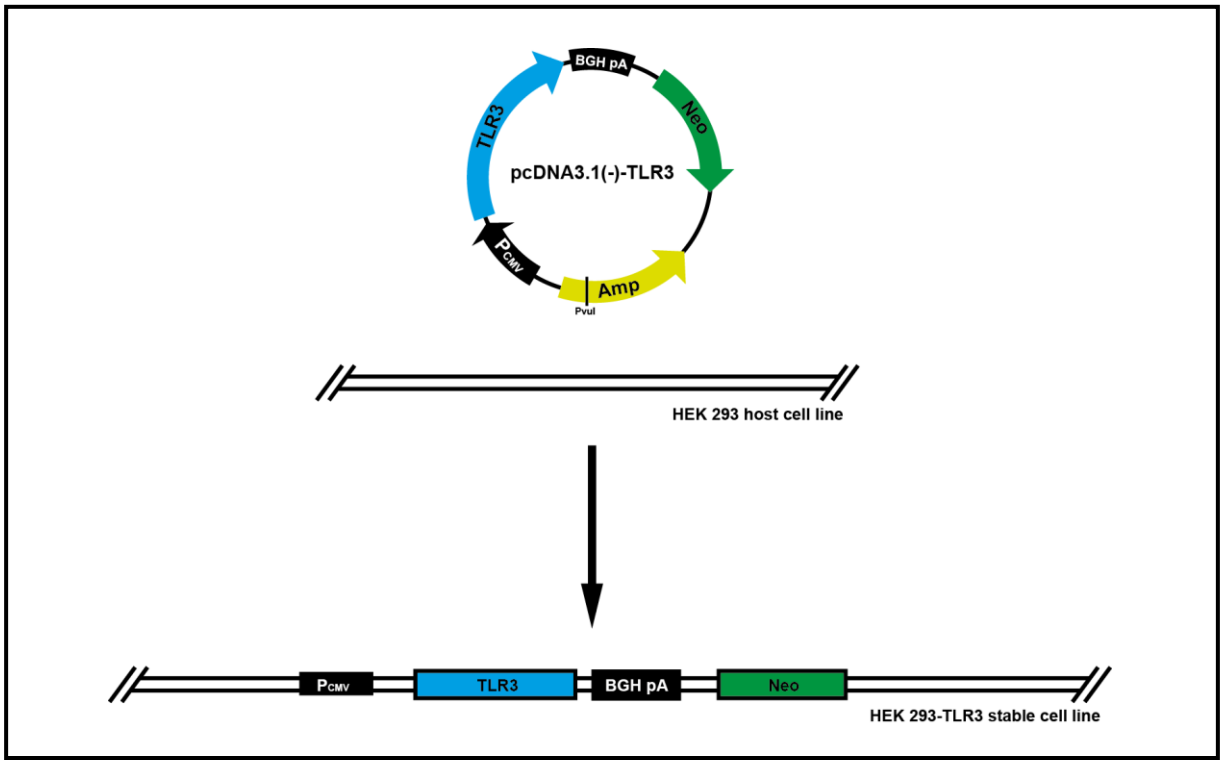

Fig. 26: Schematic representation of the generation of HEK 293-TLR3 stable cell line. pcDNA3.1(-)-TLR3 was first linearized using PvuI, purified, and then transfected into the HEK 293 cell line. 48 h after transfection, the positive control of the transfection has been observed. Growth media DMEM has been removed and fresh growth media containing Geneticin has been added to the cells for the selection of the cells containing gene of interest. After about two weeks, some clones can be observed under the microscope; the clones have been picked and transferred into a 24 well with selection media. Genomic DNA was isolated from the clone and the expression cassette was amplified using PCR.

After 48 hours of transfection, the cells are trypsinized and transferred to a 10 cm plate. Geneticin is added to the growth media to select the cells containing TLR3 and maximize the percentage of the cell population containing TLR3. Once the cell colonies are formed, they are picked and transferred to the 24-well plate. When the cells are 70% confluent in a 24-well plate, they are moved to a 6-well plate. Upon having sufficient cells, the cells are harvested and the genomic DNA of different colonies is isolated. Total expression cassette of the TLR system is amplified using PCR. The PCR products are isolated from the gel and sequenced using Sanger sequencing.

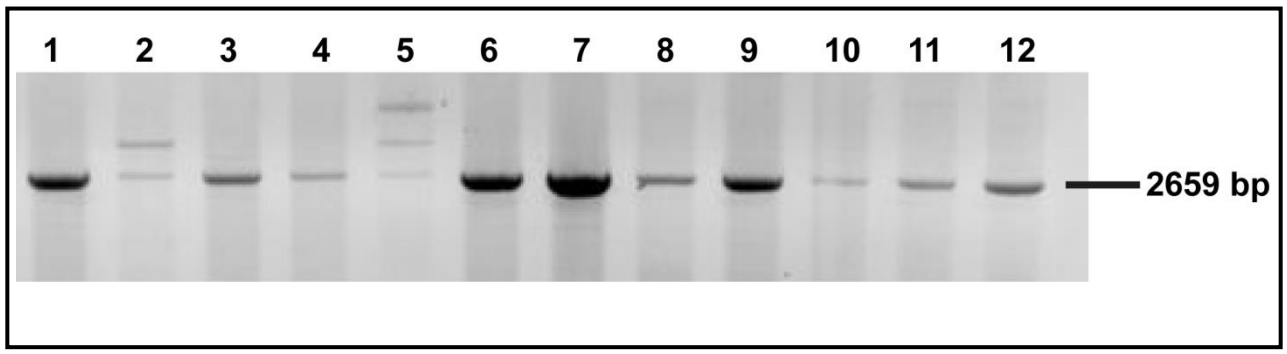

Fig. 27: Gel analysis of TLR expression cassette with CMV promoter and BGH polyA signal. Expression cassette of the TLR3 system including CMV promoter and BGH polyA was amplified using PCR, the PCR product was visualized using gel electrophoresis. The PCR product is 2695 bp. 1-12: clone number.

3.9 Activity test of CRISPR-Cas9 targets in HEK-TLR3 stable cell line

To examine the efficiency of the six targets (T3-T8) at the genomic level, the respective px459 vectors were transfected into the HEK293-TLR3 stable cell line and the fluorescing cells were counted by FACS. By measuring the BFP positive cells, up to 27% NHEJ events were observed using px459-T7, which binds within the 5’ start region of the TLR3 system and T5, which binds close to the stop codon. On the other hand, targets T3, T4, T6, and T8 show much lower NHEJ rates in the range of 5-7%.

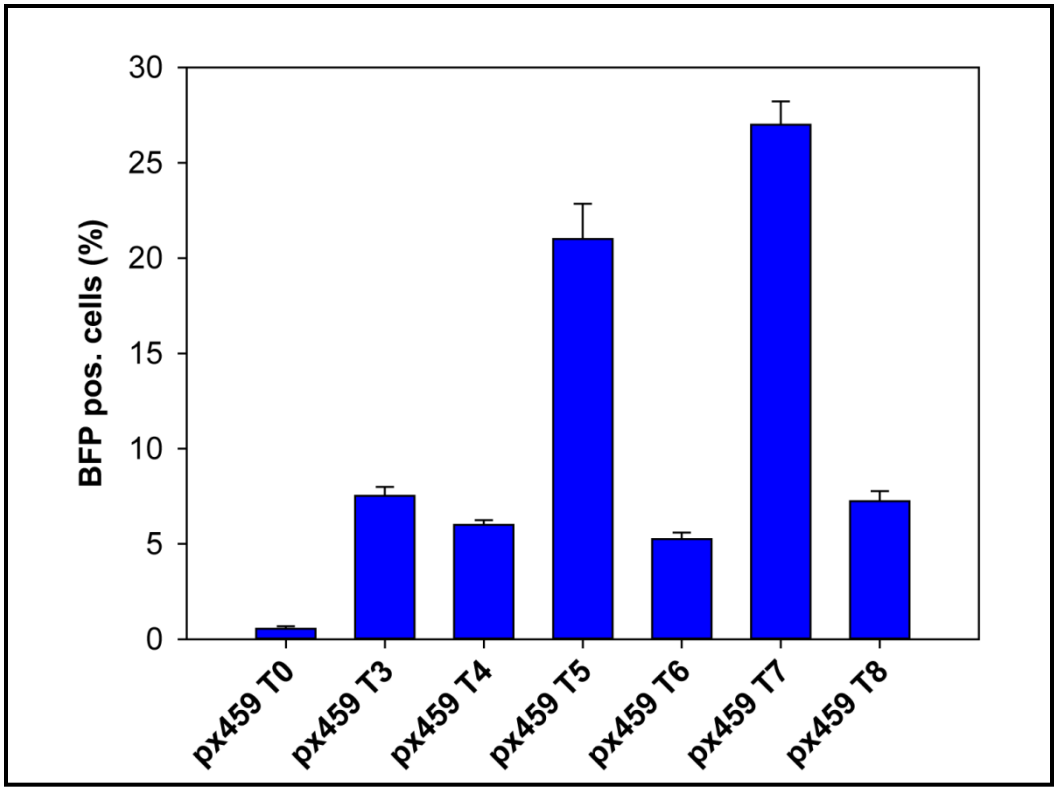

Fig. 28: Activity test of px459 CRISPR-Cas targets in HEK-TLR3 stable cell line. 1 µg nucleases were transfected into the HEK-TLR3 stable cell line. 72 h after transfection, the cells were harvested and BFP positive cells were counted using flow cytometry.

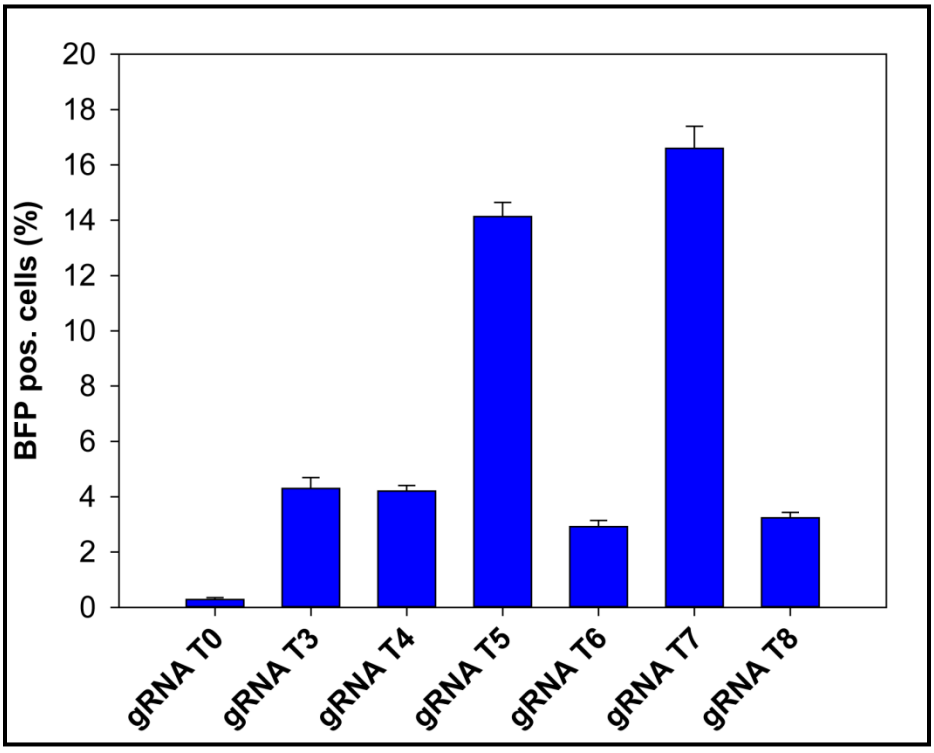

Fig. 29: Activity test of gRNA cloning vector and hCas9 expression vector in HEK-TLR3 stable cell line. 500 ng gRNA cloning vector and 500 ng hCas9 expression vector were cotransfected into the HEK-TLR3 stable cell line. 72 h after transfection, the cells were harvested and BFP positive cells were counted using flow cytometry.

3.10 HDR events using different donor templates

Since only T3 or T4 can be used in subsequent experiments and HDR levels for these two sites were rather low, the puromycine gene within the px459 plasmid was exchanged by a red fluorescent protein (mRFP, px459-mRFP) and cells were sorted prior to the DNA repair activity based signaling (blue or green signals) for red fluorescent cells, thus enriching the population of successfully transfected cells with the plasmids of interest.

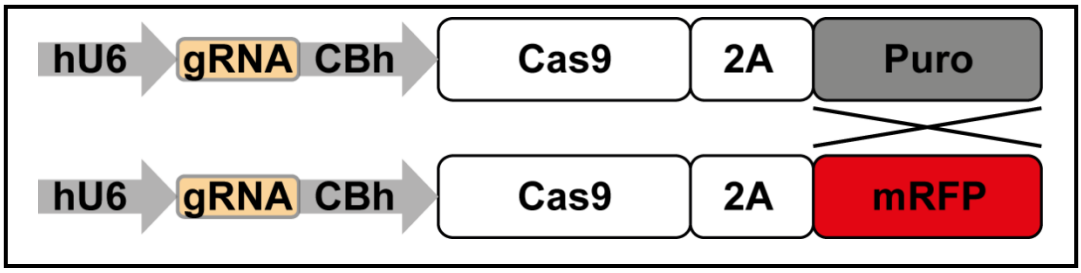

Fig. 30: px459-mRFP variant of CRISPR-Cas9 system. Puromycine gene within the px459 plasmid has been exchanged by a red fluorescent protein (mRFP) for the enrichment of the Cas9 positive cells.

Upon co-transfection of the px459-mRFP vector and the different template versions, FACS based quantification of NHEJ and HDR was performed and different template versions compared. Most importantly, the use of different donor templates had an effect on the repair outcomes for both, HDR and NHEJ, with rates ranging from 0.2% to 2.1%, and 5% to 12%, respectively.

The linearized plasmid with short 5' backbone overhang (HindIII lin. Pla.) was compared to the other three template versions in three different experiments, proving the reproducibility of the results **(Fig. 31a, c, e)**. While the absolute values varied depending on the transfection efficiency in each experiment, the overall outcome that the RS37 based template resulted in highest levels of HDR, followed by RS55, and RS73 revealing the lowest HDR activity, was constantly reproducible. In contrast, uncut plasmid DNA template RS55 showed highest HDR activity **(Fig. 31a)**, and for PCR product and the linearized plasmid with the long 5' backbone overhang (DraIII lin. Pla.) the template version RS73 revealed highest HDR activity **(Fig. 31c, e)**. Notably, linearized plasmid with short 5' backbone overhang and short 5' homology arm resulted in the highest HDR rates observed compared to any other templates **(Fig. 31a, c, e)**.

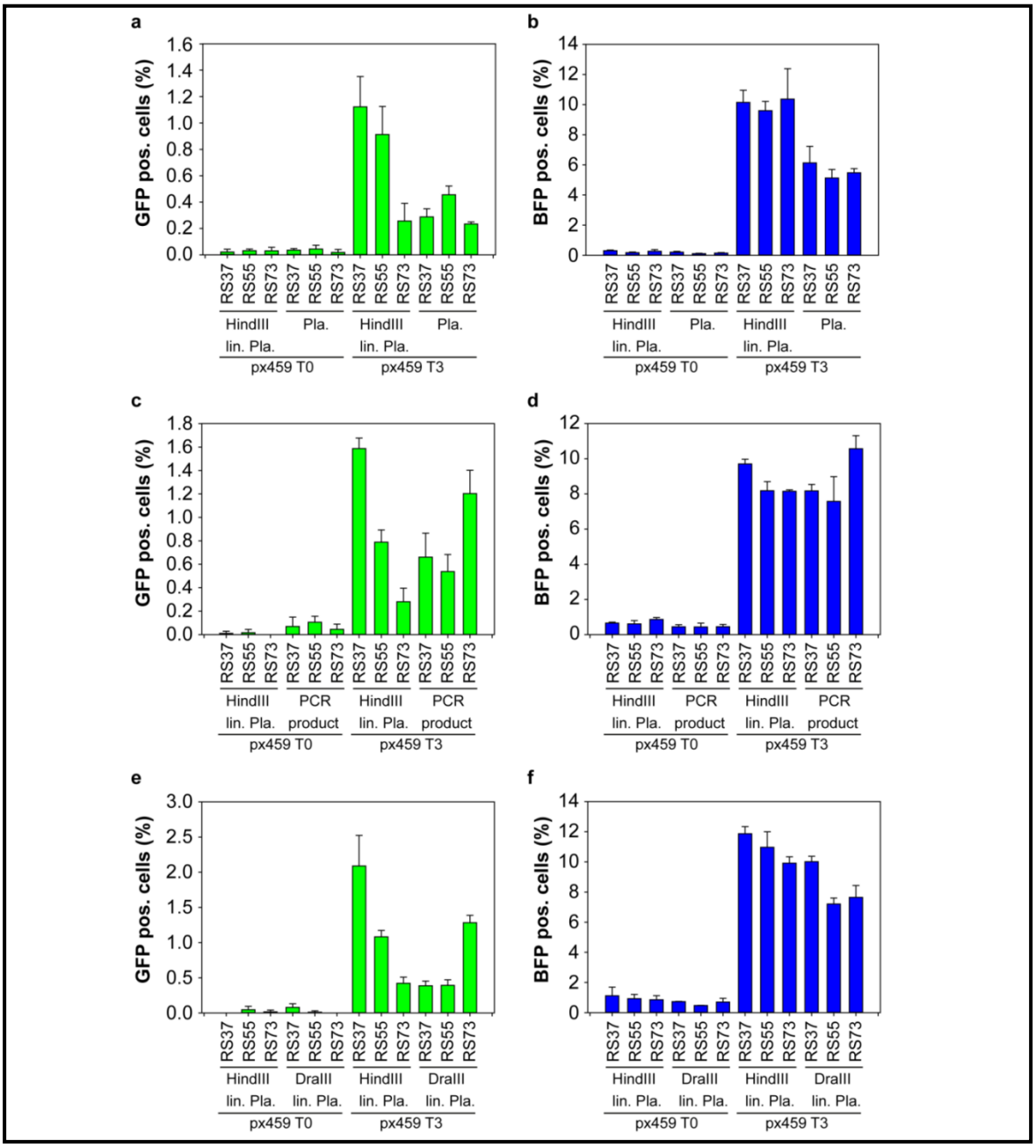

Fig. 31: Optimizing HDR events using different donor templates. Cas9-mRFP targets and donor templates were coexpressed in the HEK-TLR3 cell line. 72 h after transfection, flow cytometric analysis of mRFP+ gated cells displayed BFP+ (b, d, f, NHEJ repair) cells and GFP+ (a, c, e, HDR repair) cells. The graphs represent triplicate data from three independent experiments. T0, px459-mRFP without guide RNA. RS, repair substrate. Pla, plasmid. lin, linearized. Compared to HindIII RS37, significant differences ($p<0.005$) have been observed by the template variants HindIII RS73, Pla. RS37, Pla. RS73, PCR RS37, PCR RS55, DraIII RS37 and DraIII RS 55.

3.11 Inhibition of NHEJ key proteins

Besides the optimization of the donor templates, DNA ligase IV inhibitor Scr7 (purchased from Apexbio), and siRNA Ku80 (purchased from Origene) have been used to further optimize the HDR rate. Scr7 has been added 4 h after the transfection, while siKu80 has been cotransfected with the nucleases and the px459-T3.72 h after transfection, flow cytometric analysis of mRFP+ gated cells displayed BFP+ (NHEJ repair) cells and GFP+ (HDR repair) cells. DMSO was used as neg. control for the Scr7 inhibitor, in which Scr7 is solved. A universal scrambled siRNA control (purchased from Origene) has been used as neg. control for siKu80. A decrease of NHEJ signal and increase of HDR signal can be observed after the treatment of Scr7 and siKu80 **(Fig. 32)**.

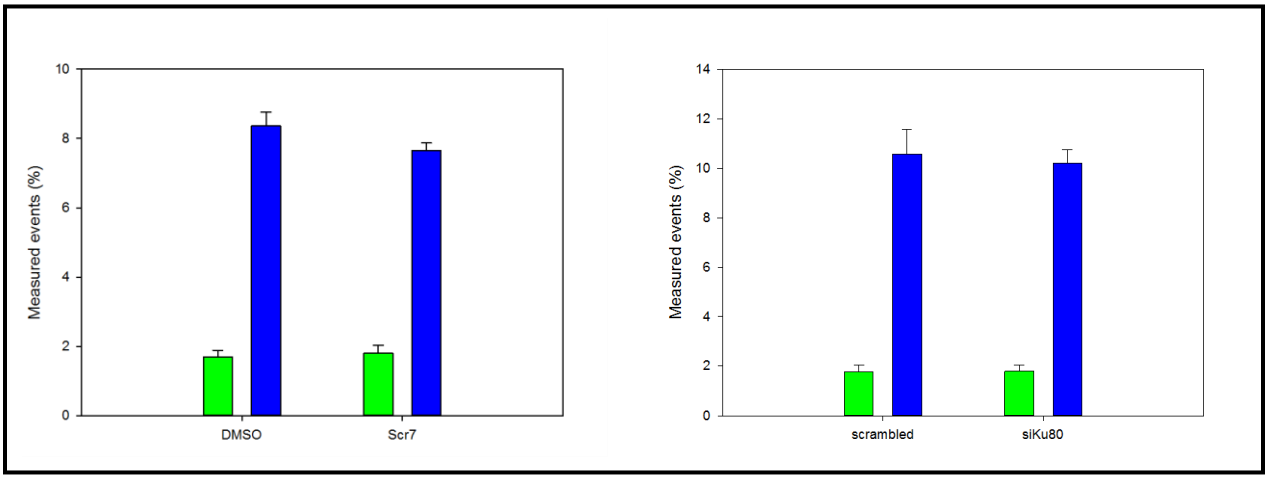

Fig. 32: Optimizing the HDR events using NHEJ key protein inhibitor. 10 mM Scr7 has been added to the media 4 h after the transfection and 10 nM siKu80 has been cotransfected with the nuclease px459-T3 and the template variant HindIII lin. Pla. RS37. Flow cytometric analysis of mRFP+ gated cells displayed BFP+ (NHEJ repair) cells and GFP+ (HDR repair) cells. The graphs represent triplicate data from three independent experiments. RS, repair substrate. Pla, plasmid. lin, linearized. Neg, negative. The increases of HDR and decreases of NHEJ observed are not significant ($p>0.005$).

4. Discussion

In therapeutic genome editing applications it is crucial to bias repair outcomes towards high fidelity homology directed repair (HDR) and to avoid error prone nonhomologous end-joining (NHEJ). In this study, the impact of different repair templates on the frequency of HDR and NHEJ was analyzed in a well-defined cell culture system. The study demonstrated that the choice of the guide RNA target site has an influence on the activity of DSB repair as measured by NHEJ activity and that target sites on the active strand result in higher DSB repair activity compared to sites on the inactive DNA strand. More importantly, it was shown that the choice of the template DNA has tremendous effects on the outcome of HDR rates, revealing important information for the subsequent preparation of genome editing approaches to treat disease-causing mutations.

4.1 The traffic light reporter system

Quantitative measurement of both HDR and NHEJ activity is an important aspect of evaluating the repair efficiency for further therapeutic genome editing application. The insertions and deletions resulting from error prone NHEJ repair are usually small and can range in size from a few to tens of nucleotides (Hendel, Fine et al. 2015). To detect these repair outcomes, different methods have been developed, such as restriction-fragment length polymorphism (RFLP) assay, SURVEYOR assay, Sanger sequencing via cloning of the DNA fragments, next generation sequencing (NGS), and some reporter-based quantification methods.

The RFLP assay normally relies on the restriction enzyme digestion of the DNA sequence changes after HDR mediated introduction of the specific recognition site, while the SURVEYOR assay uses CEL-I nuclease or T7 endonuclease I (T7E1) to digest the mismatch sequence resulting from insertions and deletions of NHEJ repair (Pourzand and Cerutti 1993, Qiu, Shandilya et al. 2004). Both RFLP and SURVEYOR assays are gel-based mutation detection

assays, which are rapid and cost-effective but the sensitivity is relative low. For the comparison of HDR activity changes due to different templates used in this study, with differences estimated to be below 1%, these methods are not suitable.

Table 16: Comparison of assays for quantifying genome editing outcomes

Methods to quantify repair outcomes	Quantification of HDR	Quantification of NHEJ	Cost	Sensitivity	Throughput	Ref.
RFLP assay	Yes	No	Low	Low	Low	(Pourzand and Cerutti 1993)
SURVEYOR assay	No	Yes	Low	Low	Low	(Qiu, Shandilya et al. 2004)
Sanger sequencing (Single cell clone analysis)	Yes	Yes	Low	High	Low	(Ran, Hsu et al. 2013)
NGS	Yes	Yes	High	High	High	(Ran, Hsu et al. 2013)
DR-GFP reporter system	Yes	No	Low	High	High	(Pierce, Johnson et al. 1999)
Integration fluorescent gene	Yes	No	Low	High	High	(Moehle, Rock et al. 2007)
BFP reporter system	Yes	Yes	Low	High	High	(Richardson, Ray et al. 2016)
TLR system	Yes	Yes	Low	High	High	(Certo, Ryu et al. 2011)
TLR3 system	Yes	Yes	Low	High	High	This study

Targeted genome modifications can also be measured by either Sanger sequencing or next generation sequencing (NGS). In Sanger sequencing-based assays, the modified DNA sequences are first isolated from the cells and amplified using PCR. The amplicons are subcloned into a plasmid vector such as TOPO or pUC19 for transformation, and individual colonies can be sequenced by Sanger sequencing for further sequence analysis. Although this sequencing method can give more detailed information about the modified events in the genome, the rather small number of readouts could be a limiting factor to be considered. Recently, NGS has become a popular method to detect repair outcomes. NGS can detect both HDR and NHEJ simultaneously with high sensitivity reliably to 0.01% (Hendel, Fine et al. 2015). In contract to Sanger sequencing, NGS is more efficient to quantify a large number of samples or target sites with high throughput. It should be noted that these sequencing-based assays rely on PCR amplicons and large deletions and insertions that span beyond the boundaries of the PCR amplicon are potentially less likely amplified and therefore cannot be detected.

Several flow-cytometric based fluorescent reporter systems are also often used to detect the repair outcomes. Addition of a fluorescent gene to a defined locus in human cells using donor DNA containing locus-specific homology arms and the GFP ORF together with a ZFN demonstrated 1.4% GFP positive cells, indicating the successful HDR-mediated DNA repair (Moehle, Rock et al. 2007). However, this assay cannot quantify the NHEJ repair simultaneously and thus the nuclease activity must be quantified using other method. Another approach to estimate the efficiency of HR and NHEJ is the BFP reporter system, which is based on the conversion of BFP reporter to GFP via a three-nucleotide mutation (Richardson, Ray et al. 2016). It enables a rapid observation of both HDR and NHEJ simultaneously. Although this proof-of-concept reporter system is highly useful for optimizing DNA repair pathway choices for small sequence changes such as SNP mutations, it is not suitable to be used as a model for large sequence changes.

To overcome these limitations, traffic light reporter system (TLR) has been used in this study (Certo, Ryu et al. 2011). The traffic light reporter system is able to directly measure the efficiency and competition between DNA repair pathway choices at individual DNA breaks. It contains a non-functional green fluorescent protein (GFP) gene and a second blue fluorescent

protein (BFP) gene in a reading frame shifted by 2 bp. 38 bp sequence changes within the mutated GFP sequence is of a comparable size to indel mutations discovered in human genomes (Mullaney, Mills et al. 2010). Upon repair of the double-strand break (DSB) close to the stop codon within the 38 bp mutated GFP sequence, differing fluorescent signals will appear depending on whether homology directed repair (HDR) or non-homologous end-joining (NHEJ) takes place.

Of note, BFP signal is generated by +3 frameshifts, but not all indels will shift the reading frame into that of BFP, thus +3 frameshifts represents about 1/3 of all the mutagenic events (Certo, Ryu et al. 2011). In contrast to BFP signal, GFP signal was generated through the precise repair of the GFP cDNA sequence and the normal genetic function is restored by the exogenously provided donor template, which shows all HDR events. In this work, all measured cells have been represented in the graphic but not the repair events, since the stop codon could be created randomly after DSB, and through that a minority -2 frameshifts will not express BFP. Therefore, the measured BFP signal is not exact 1/3 of the NHEJ events. However, since we compare data from different templates using the same TLR3 system, it does not matter, what the exact number of NHEJ events is, but the comparison among the different settings.

An important limitation of this work is the fact that FACS is a cell based detection method. But the HEK-TLR3 cell line was generated via random integration resulting in an unknown number of copies of the TLR3 reporter sequence in each individual cell. It could be enough to detect the fluorescent signal even if one TLR copy is repaired among the unknown copies in the whole genome. Nonetheless, to some extent this is a minor consideration for this initial proof-of-concept work, since the comparison of the donor templates and the nuclease activity is under the same condition.

4.2 Cleavage at active or inactive strand

In this study, six different target sites were identified upstream or downstream of the stop codon within the mutated GFP sequence of the TLR3 system containing the initiating 5'G and the 3'PAM (NGG) sequence, among which T3 and T4 cleavage sites are inside the mutated GFP sequence. Furthermore, T5 and T7 are located on the transcriptionally activate strand of the TLR3 cDNA sequence, while T3, T4, T6 and T8 are located on the transcriptionally inactive strand of the cDNA sequence. After the quantification of the BFP (NHEJ) positive cells, all six targets showed nuclease activity, among which T5 and T7 resulted in higher NHEJ signals and the other 4 target sites resulted in lower NHEJ signals.

Interestingly, the two targets with higher activity are all located on the transcriptionally active strand of the DNA, while the four sites with lower activity are located on the transcriptionally inactive strand of the DNA, indicating that NHEJ rates are dependent on the location of the target site being on the transcriptionally active or inactive DNA strand. Similar results have also been observed by Cong and colleagues (Cong, Ran et al. 2013). 16 target sites based on *Streptococcus pyogenes* type II CRISPR and *Streptococcus thermophilus* CRISPR1 loci have been designed against three different genes in human and mouse genomes, most of the targets located on the active plus strand have shown higher indel rates than the targets located on the inactive minus strand (Cong, Ran et al. 2013).

Recently, the interaction of Cas9 with target DNA has been studied by Richardson and colleagues (Richardson, Ray et al. 2016). Understanding the mechanisms of Cas9-DNA interaction could suggest strategies to design targets for effective cleavage. They have shown that dissociation of Cas9 from double-stranded DNA (dsDNA) substrates is asymmetric and, before complete dissociation, Cas9 releases the 3' end of the PAM distal nontarget strand while holding onto other three ends of the target DNA (white crossed circles). Our observation that gRNAs targeting the transcriptionally active strand resulted in higher NHEJ signal than gRNAs targeting the transcriptionally inactive strand suggests that small indels are more likely to be introduced at the

DNA strand that Cas9 releases, and sequence changes on the active plus strand have more influence on the NHEJ repair outcome than that on the inactive minus strand.

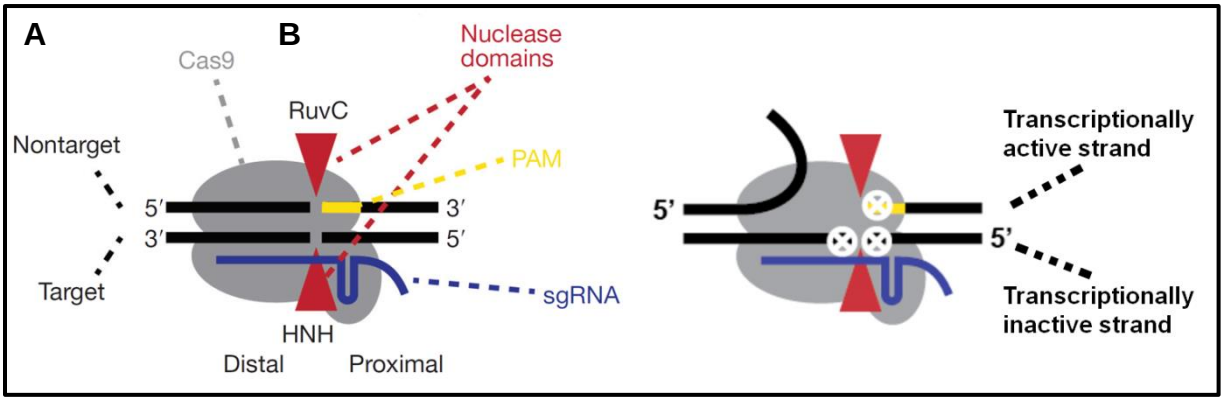

Fig. 33: Interaction of Cas9 with target DNA. (A) Cas9 (gray) binds stably to substrate DNA (black). RuvC and HNH nuclease domains cut the nontarget and target strands, respectively (red triangles). **(B)** After cleavage, PAM distal non-target strand is released from the Cas9-DNA complex before complete dissociation. Modified after (Richardson, Ray et al. 2016).

The position of the gRNA target sequence being on the transcriptional active or inactive strand has also been analyzed using nicking enzymes and different types of donor template, and the authors observed increased DSB repair activity when the site was positioned on the active strand (Davis 2014). Although the experimental setups are something different to this study, it suggests that designing target sites against the transcriptionally active strand could increase NHEJ activity.

4.3 Template DNA

The efficiency of HDR is determined by many factors, of which the donor template is considered to be among of the most important ones (Richardson, Ray et al. 2016). Linearized or double-stranded DNA plasmid sequences, as well as single-stranded DNA oligonucleotides, are often used as donor template for large or small sequence changes in the presence of target cleavage

(Carroll and Beumer 2014). The length of the homology arm also plays an important role in increasing HDR rate since the efficiency of recombination increases as the length of homology arms increases (Li, Wang et al. 2014). Here, 1 kb total homology sequence has been chosen for the recombination either without CMV promoter sequence or without intact GFP sequence. The templates have been generated as double-stranded DNA, PCR product, and linearized plasmid at two different backbone sites.

The observation that linearized plasmid is more effective than plasmid donor suggests that circular DNAs can be broken randomly at undesired site in the homology arm. In contrast to results from the current study in HEK cells, it was observed that plasmid donor is more effective than linearized donor in Drosophila embryos probably due to the degradation of the linear DNA by exonucleases or conversion of long concatemers (Carroll 2011). The fate of the plasmid donor and linearized donor could also be determined by the delivery system since the duration of the donor in the cytoplasm increases the degradation rates.

The issue of linearized plasmid vs. PCR product is also an interesting one. Upon generation of the plasmid DNA, post-translational modification of the bacteria could be a limiting factor for homology directed repair. Another possible reason of the higher HDR efficiency of the linearized plasmid is the protection of the backbone sequence from degradation of the template sequence, which is not the case for PCR products.

Moreover, the plasmids were linearized at 5’ and 3’ end for the generation of the backbone overhang at different sites. Surprisingly, the cut side of the plasmid has resulted indifferent HDR and NHEJ rates. It is possible that the overhangs resulted from HindIII and DraIII have different affinity to the DNA ends generated from Cas9 protein, so that the random integration of the templates into the TLR3 cut site could be different, which in turn influences the NHEJ efficiency. Our observation that linearized RS37 showed the best HDR rate suggests that 3’ end degradation takes places at the break, while the overhang of the backbone protects the degradation of the 3’ end.

Furthermore, we have also tried to use ssONGs (100 bp) to stimulate HDR events in our setup, but it was not successful, probably due to the large sequence changes (38 bp) inside the TLR3 system (data not shown). This study showed the importance of choosing the right DNA template molecule as one important aspect in the development of future gene therapeutic approaches.

Another key issue in HDR repair is the DNA conversion tract length (DCTL). DCTL shows how much donor sequence is transferred to the target and how far synthesis goes during D loop formation between the broken DNA and the intact homologous donor sequence. In mammalian cells, an early study has examined DCTL in 80 recombinants and has observed relatively short DCTL, with 80% of the recombinants having tracts of 58 bp or less (Elliott, Richardson et al. 1998). A steep decline in the amount of conversion around the double strand break site and up to 511 bp of gene conversion tracts has been identified in this study. These results suggest that the broken ends are protected from extensive degradation prior to or during recombination.

Similar results have been observed in another study, in which DCTL has been tested in Drosophila in two different experimental protocols (Beumer, Trautman et al. 2013). After induction of ZFN and different donor templates, conversion tracts fell by half about 500 bp on either side of the targeted cut site with symmetric distribution and still 50% conversion frequency was observed 3 kb or more from the ZFN cut site in the embryo injection protocol. Although the tendencies of these two studies are more or less similar, the absolute values of the DCTL are totally different, probably due to the different cell types and different length of donor template **(Fig. 34)**. Elliott et al has used donor template of 745 bp, while Beumer et al has used 4.2 and 7.5 kb donor.

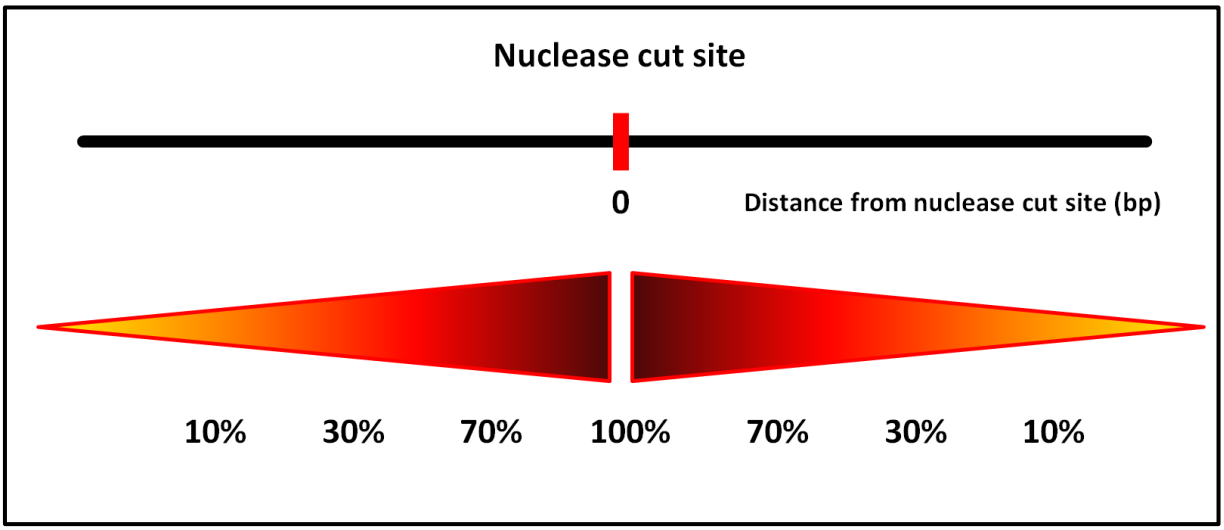

Fig. 34: DNA conversion tract length during homologous recombination. DCTL is dependent on the distance from nuclease cut site and the frequency is dependent on other related factors, such as cell type and DNA donor template. The value at the cut site is necessarily 100%.

In this study, CRISPR-Cas9-T3 is directly located within the 38 bp mutated sequence of the TLR3 system for the measurement of HDR events, which indicates a higher rate of conversion tracts compared to the previous studies although it could not be directly measured. By using other targets such as CRISPR-Cas9-T7 with a distance of about 100 bp away from the mutated GFP sequence, it is also possible to measure the DCTL in the TLR3 system.

4.4 Approaches to increase HDR

Beside the optimization of the DNA donor templates, several approaches have been developed to avoid NHEJ and bias repair towards error free pathways. Regulation of expression of key repair pathway proteins, such as inhibiting NHEJ factors by siRNA technology or specific blockers, or overexpression of HDR key proteins is another way to enhance HDR (Yanik, Muller et al. 2016). Generally, siRNA technology can be applied to all NHEJ proteins such as the KU70/80 complex, DNA-PKcs, Artemis and DNA Ligase IV/XRCC4, while inhibitors have been only reported for DNA-PKcs and DNA Ligase IV until recently (Srivastava and Raghavan 2015) **(Table 17)**.

Table 17: NHEJ inhibitors

Gene	Inhibitor	Ref.
DNA Ligase IV	SCR7 L189	(Srivastava, Nambiar et al. 2012) (Chen, Zhong et al. 2008)
DNA-PKcs	wortmannin NU7026 LY294002	(Oliveira, Castro et al. 2002) (Willmore, de Caux et al. 2004) (Rosenzweig, Youmell et al. 1997)

DNA-PKcs and KU70/80 complex play an important role in the initial step of NHEJ. Targeting proteins involved in the early stage of NHEJ could divert the repair pathway to alternative NHEJ or HDR. In contrast, inhibition of the proteins at the final stage of NHEJ, such as DNA LigaseIV could lead to accumulation of DNA breaks, which can be lethal to the cells. However, it has been reported that HDR can be significantly increased by using SCR7, a DNA ligase IV inhibitor (Chu, Weber et al. 2015, Maruyama, Dougan et al. 2015). Chu et al has reported that the efficiency of HDR was increased from 5% to 8-14% in the presence of single shRNAs against KU70, KU80 or DNA ligase IV, to 25% by using shRNAs against KU70 and DNA ligase IV or SCR7 (Chu, Weber et al. 2015). Up to 19-fold increased HDR efficiency has been observed by Maruyama and colleagues by using SCR7 in mammalian cell lines and in mice for four genes (Maruyama, Dougan et al. 2015).

Based on these successful previous studies, SCR7 and siRNA against KU80 have been chosen to increase HDR in this study. Similar to these two studies, decreased NHEJ activity has been observed. Unfortunately, HDR activity did not increased significantly. Compared to the previous successful studies, concentration, incubation time, transfection method and cell type could be the limiting factors resulting in different outcomes. Further optimization is needed to increase the HDR activity.

These approaches, although they seem to be highly useful in some contexts, may be undesirable during the *in vivo* therapeutic application because the molecules are toxic and can alter the

cellular response to DNA damage at other non-target sites in the genome and lead to tumor formation (Richardson, Ray et al. 2016). Further understanding of the mechanisms and optimizing the toxicity are necessary for the clinical application in the future.

4.5 Outlook

This study demonstrated the importance of designing CRISPR-Cas9 targets and DNA donor templates for the development of HDR mediated therapeutic genome editing approach. The limitation of this study is that the off-target toxicity of the CRISPR-Cas9 targets was not analyzed. Now, it is of little importance for this proof-of-principle work, but for the long term therapeutic application, the off-target analysis needs to be completed. Random integration of introduced DNA such as CRISPR-Cas9 expression vector and DNA donor template vector is another issue that should be considered.

Once these approaches have been established in the cell culture, they will be applied *in vivo* in the TLR mouse model. The TLR mouse model has been generated by Stieger and colleagues for the development of a treatment strategy for X-linked retinitis pigmentosa, of which TLR is located adjacent to the mutational hotspot ORF15 of the retinitis pigmentosa GTPase regulator (RPGR) gene on the X-chromosome (Stieger, unpublished data). Repair of the TLR system in the RPGR gene suggests that the disease-causing mutations in the ORF15 could also be repaired. Subsequently, patients suffering from XLRP might be treated by this technique in the long term future.

Summary

The CRISPR-Cas technology enables rapid and precise genome editing at any desired genomic position in almost all cells and organisms. For therapeutic application, it is crucial to bias repair outcomes towards high fidelity homology directed repair (HDR) and to avoid error prone nonhomologous end-joining (NHEJ). In this study, the impact of different repair templates on the frequency of homology directed repair (HDR) and non-homologous end-joining (NHEJ) has been analyzed.

A stable HEK293 cell line expressing TLR3 was used to quantify HDR and NHEJ events. The modified TLR system (TLR3) comprises a bicistronic expression system of a non-functional green fluorescent protein (GFP) gene, followed by a self-cleaving T2A peptide and a second blue fluorescent protein (BFP) gene in a reading frame shifted by 2 bp. A stable HEK293 cell line expressing TLR3 was generated by transfecting a linearized pcDNA3.1(-)-TLR3 plasmid followed by neomycin selection. Donor templates of 1000 bp length containing the corrected GFP sequence were generated as circular plasmid, linearized plasmid with long 3’ or 5’ backbone overhang, or as PCR product. The sequence to be corrected was either centrally located (RS55), with a shorter 5’ homologous region (RS37), or a shorter 3’ homologous region (RS73). Six different CRISPR-Cas9 target sites were identified upstream or downstream of the stop codon within the GFP sequence containing the initiating 5’G and the 3’PAM (NGG).DNA repair activity was measured by FACS.

Guide RNAs targeting the active strand (T5, T7) showed higher NHEJ frequencies compared to guide RNAs targeting the inactive strand. HDR activity was highest when using the linearized plasmid with the short 5’ backbone overhang and the RS37 design, followed by the PCR product or the linearized plasmid with the long 5’ backbone overhang, both with RS73 design. Circular plasmid was least efficient in generating HDR events.The effect of the different repair templates on NHEJ frequencies was marginal. The results demonstrate the importance of the design of the

guide RNA and template DNA on the frequency of DNA repair events and thus, ultimately on the outcome of treatment approach using HDR.

Zusammenfassung

Durch die CRISPR-Cas Technologie ist es möglich, schnelle und genaue Veränderungen in jeder gewünschten Gensequenz, in nahezu jedem Zelltyp und Organismus durchzuführen. Wird die Technologie zu Therapiezwecken genutzt, ist es wichtig, die Ergebnisse der Reparatur in Richtung HDR zu lenken und das fehlerhafte NHEJ zu vermeiden. In dieser Studie wurde der Einfluss verschiedener Reparatur-*Templates* auf die HDR- und NHEJ-Rate analysiert.

Eine stabile HEK293 Zelllinie, die das TLR3 exprimiert, wurde verwendet um die HDR und NHEJ Events zu quantifizieren. Das modifizierte TLR System (TLR3) beinhaltet ein bicistronisches Expressionssystem. Dieses besteht aus einem nicht funktionellen GFP-Gen (grünes Fluoreszenz Protein), gefolgt von einem selbst spaltenden T2A Peptid und einem BFP-Gen (blaues Fluoreszenz Protein) in einem Leserahmen, der um zwei Basenpaare verschoben ist. Eine stabile HEK293 Zelllinie, die das TLR3 exprimiert, wurde durch Transfektion des linearisierten pcDNA3.1(-)-TLR3 Plasmids mit anschließender Neomycin Selektion generiert. Die *Donor-Templates* mit 1000 Basenpaaren Länge, die die korrigierte GFP Sequenz beinhalten, wurden in Form des zirkulären Plasmids, des linearisierten Plasmids mit langem 3‘ oder 5‘ Backbone Überhang, oder des PCR Produkts verwendet. Die zu korrigierende Sequenz liegt entweder zentral (RS55), besitzt eine kürzere 5‘ homologe Region (RS37), oder eine kürzere 3‘ homologe Region (RS73). Sechs verschiedene CRISPR-Cas9 *targetsites* wurden *upstream* oder *downstream* des Stopp-Codons innerhalb der GFP Sequenz, die das initiale 5’G und die 3‘ PAM (NGG) beinhaltet, identifiziert. Die DNA Reparaturaktivität wurde durch FACS gemessen.

Die Guide RNAs, die den aktiven Strang adressieren (T5, T7), zeigen eine höhere NHEJ Rate im Vergleich zu den guide RNAs, die den inaktiven Strang adressieren. Die höchste HDR Aktivität konnte mit Hilfe des linearisierten Plasmids mit kurzem 5‘ Backbone Überhang und RS37 Design erzielt werden, gefolgt von dem PCR Produkt oder dem linearisierten Plasmid mit langem 5‘ Backbone Überhang, beide mit RS73 Design. Das zirkuläre Plasmid erzeugte am wenigsten

HDR Events. Der Effekt der verschiedenen Reparatur-Templates auf die NHEJ Rate ist marginal. Die Ergebnisse zeigen den Einfluss des Designs der guide RNA und der Template DNA auf die Rate der DNA Reparatur-Events und damit letztlich auf den Erfolg des Therapie-Ansatzes durch das Nutzen von HDR.

Literatures

Amitai, G. and R. Sorek (2016). "CRISPR-Cas adaptation: insights into the mechanism of action." Nat Rev Microbiol**14**(2): 67-76.

Bainbridge, J. W., A. J. Smith, S. S. Barker, S. Robbie, R. Henderson, K. Balaggan, A. Viswanathan, G. E. Holder, A. Stockman, N. Tyler, S. Petersen-Jones, S. S. Bhattacharya, A. J. Thrasher, F. W. Fitzke, B. J. Carter, G. S. Rubin, A. T. Moore and R. R. Ali (2008). "Effect of gene therapy on visual function in Leber's congenital amaurosis." N Engl J Med**358**(21): 2231-2239.

Barrangou, R., C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. A. Romero and P. Horvath (2007). "CRISPR provides acquired resistance against viruses in prokaryotes." Science**315**(5819): 1709-1712.

Beumer, K. J., J. K. Trautman, K. Mukherjee and D. Carroll (2013). "Donor DNA Utilization during Gene Targeting with Zinc-finger Nucleases." G3 (Bethesda).

Bitinaite, J., D. A. Wah, A. K. Aggarwal and I. Schildkraut (1998). "FokI dimerization is required for DNA cleavage." Proc Natl Acad Sci U S A**95**(18): 10570-10575.

Blaese, R. M., K. W. Culver, A. D. Miller, C. S. Carter, T. Fleisher, M. Clerici, G. Shearer, L. Chang, Y. Chiang, P. Tolstoshev, J. J. Greenblatt, S. A. Rosenberg, H. Klein, M. Berger, C. A. Mullen, W. J. Ramsey, L. Muul, R. A. Morgan and W. F. Anderson (1995). "T lymphocyte-directed gene therapy for ADA- SCID: initial trial results after 4 years." Science**270**(5235): 475-480.

Boch, J. and U. Bonas (2010). "Xanthomonas AvrBs3 family-type III effectors: discovery and function." Annu Rev Phytopathol**48**: 419-436.

Bohgaki, M., T. Bohgaki, S. El Ghamrasni, T. Srikumar, G. Maire, S. Panier, A. Fradet-Turcotte, G. S. Stewart, B. Raught, A. Hakem and R. Hakem (2013). "RNF168 ubiquitylates 53BP1 and controls its response to DNA double-strand breaks." Proc Natl Acad Sci U S A**110**(52): 20982-20987.

Bonas, U., R. E. Stall and B. Staskawicz (1989). "Genetic and structural characterization of the avirulence gene avrBs3 from Xanthomonas campestris pv. vesicatoria." Mol Gen Genet**218**(1): 127-136.

Bondy-Denomy, J. and A. R. Davidson (2014). "To acquire or resist: the complex biological effects of CRISPR-Cas systems." Trends Microbiol**22**(4): 218-225.

Brunzell, J. D. (1993). Familial Lipoprotein Lipase Deficiency. GeneReviews(R). R. A. Pagon, M. P. Adam, H. H. Ardinger et al. Seattle (WA).

Bryant, L. M., D. M. Christopher, A. R. Giles, C. Hinderer, J. L. Rodriguez, J. B. Smith, E. A. Traxler, J. Tycko, A. P. Wojno and J. M. Wilson (2013). "Lessons learned from the clinical development and market authorization of Glybera." Hum Gene Ther Clin Dev**24**(2): 55-64.

Carroll, D. (2011). "Genome engineering with zinc-finger nucleases." Genetics**188**(4): 773-782.

Carroll, D. and K. J. Beumer (2014). "Genome engineering with TALENs and ZFNs: repair pathways and donor design." Methods**69**(2): 137-141.

Cavazzana-Calvo, M., E. Payen, O. Negre, G. Wang, K. Hehir, F. Fusil, J. Down, M. Denaro, T. Brady, K. Westerman, R. Cavallesco, B. Gillet-Legrand, L. Caccavelli, R. Sgarra, L. Maouche-Chretien, F. Bernaudin, R. Girot, R. Dorazio, G. J. Mulder, A. Polack, A. Bank, J. Soulier, J. Larghero, N. Kabbara, B. Dalle, B. Gourmel, G. Socie, S. Chretien, N. Cartier, P. Aubourg, A. Fischer, K. Cornetta, F. Galacteros, Y. Beuzard, E. Gluckman, F. Bushman, S. Hacein-Bey-Abina and P. Leboulch (2010). "Transfusion independence and HMGA2 activation after gene therapy of human beta-thalassaemia." Nature**467**(7313): 318-322.

Ceccaldi, R., B. Rondinelli and A. D. D'Andrea (2016). "Repair Pathway Choices and Consequences at the Double-Strand Break." Trends Cell Biol**26**(1): 52-64.

Certo, M. T., B. Y. Ryu, J. E. Annis, M. Garibov, J. Jarjour, D. J. Rawlings and A. M. Scharenberg (2011). "Tracking genome engineering outcome at individual DNA breakpoints." Nat Methods**8**(8): 671-676.

Chandrasegaran, S. and D. Carroll (2016). "Origins of Programmable Nucleases for Genome Engineering." J Mol Biol**428**(5 Pt B): 963-989.

Chang, N., C. Sun, L. Gao, D. Zhu, X. Xu, X. Zhu, J. W. Xiong and J. J. Xi (2013). "Genome editing with RNA-guided Cas9 nuclease in zebrafish embryos." Cell Res**23**(4): 465-472.

Chen, B., L. A. Gilbert, B. A. Cimini, J. Schnitzbauer, W. Zhang, G. W. Li, J. Park, E. H. Blackburn, J. S. Weissman, L. S. Qi and B. Huang (2013). "Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system." Cell**155**(7): 1479-1491.

Chen, X., H. Niu, W. H. Chung, Z. Zhu, A. Papusha, E. Y. Shim, S. E. Lee, P. Sung and G. Ira (2011). "Cell cycle regulation of DNA double-strand break end resection by Cdk1-dependent Dna2 phosphorylation." Nat Struct Mol Biol**18**(9): 1015-1019.

Chen, X., S. Zhong, X. Zhu, B. Dziegielewska, T. Ellenberger, G. M. Wilson, A. D. MacKerell, Jr. and A. E. Tomkinson (2008). "Rational design of human DNA ligase inhibitors that target cellular DNA replication and repair." Cancer Res**68**(9): 3169-3177.

Chira, S., C. S. Jackson, I. Oprea, F. Ozturk, M. S. Pepper, I. Diaconu, C. Braicu, L. Z. Raduly, G. A. Calin and I. Berindan-Neagoe (2015). "Progresses towards safe and efficient gene therapy vectors." Oncotarget**6**(31): 30675-30703.

Chu, V. T., T. Weber, B. Wefers, W. Wurst, S. Sander, K. Rajewsky and R. Kuhn (2015). "Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells." Nat Biotechnol**33**(5): 543-548.

Coluccio, A., F. Miselli, A. Lombardo, A. Marconi, G. Malagoli Tagliazucchi, M. A. Goncalves, C. Pincelli, G. Maruggi, M. Del Rio, L. Naldini, F. Larcher, F. Mavilio and A. Recchia (2013). "Targeted gene addition in human epithelial stem cells by zinc-finger nuclease-mediated homologous recombination." Mol Ther**21**(9): 1695-1704.

Cong, L., F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini and F. Zhang (2013). "Multiplex genome engineering using CRISPR/Cas systems." Science**339**(6121): 819-823.

Deng, D., C. Yan, X. Pan, M. Mahfouz, J. Wang, J. K. Zhu, Y. Shi and N. Yan (2012). "Structural basis for sequence-specific recognition of DNA by TAL effectors." Science**335**(6069): 720-723.

Deriano, L. and D. B. Roth (2013). "Modernizing the nonhomologous end-joining repertoire: alternative and classical NHEJ share the stage." Annu Rev Genet**47**: 433-455.

Deyle, D. R. and D. W. Russell (2009). "Adeno-associated virus vector integration." Curr Opin Mol Ther**11**(4): 442-447.

DiCarlo, J. E., J. E. Norville, P. Mali, X. Rios, J. Aach and G. M. Church (2013). "Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems." Nucleic Acids Res**41**(7): 4336-4343.

Dickinson, D. J., J. D. Ward, D. J. Reiner and B. Goldstein (2013). "Engineering the Caenorhabditis elegans genome using Cas9-triggered homologous recombination." Nat Methods**10**(10): 1028-1034.

Elliott, B., C. Richardson, J. Winderbaum, J. A. Nickoloff and M. Jasin (1998). "Gene conversion tracts from double-strand break repair in mammalian cells." Mol Cell Biol**18**(1): 93-101.

Elsner, M., T. Terbish, A. Jorns, O. Naujok, D. Wedekind, H. J. Hedrich and S. Lenzen (2012). "Reversal of diabetes through gene therapy of diabetic rats by hepatic insulin expression via lentiviral transduction." Mol Ther**20**(5): 918-926.

Friedland, A. E., Y. B. Tzur, K. M. Esvelt, M. P. Colaiacovo, G. M. Church and J. A. Calarco (2013). "Heritable genome editing in C. elegans via a CRISPR-Cas9 system." Nat Methods**10**(8): 741-743.

Gabriel, R., A. Lombardo, A. Arens, J. C. Miller, P. Genovese, C. Kaeppel, A. Nowrouzi, C. C. Bartholomae, J. Wang, G. Friedman, M. C. Holmes, P. D. Gregory, H. Glimm, M. Schmidt, L. Naldini and C. von Kalle (2011). "An unbiased genome-wide analysis of zinc-finger nuclease specificity." Nat Biotechnol**29**(9): 816-823.

Gaudet, D., J. Methot and J. Kastelein (2012). "Gene therapy for lipoprotein lipase deficiency." Curr Opin Lipidol**23**(4): 310-320.

Genovese, P., G. Schiroli, G. Escobar, T. Di Tomaso, C. Firrito, A. Calabria, D. Moi, R. Mazzieri, C. Bonini, M. C. Holmes, P. D. Gregory, M. van der Burg, B. Gentner, E. Montini, A. Lombardo and L. Naldini (2014). "Targeted genome editing in human repopulating haematopoietic stem cells." Nature**510**(7504): 235-240.

Ghosh, A., Y. Yue, Y. Lai and D. Duan (2008). "A hybrid vector system expands adeno-associated viral vector packaging capacity in a transgene-independent manner." Mol Ther**16**(1): 124-130.

Gilbert, L. A., M. A. Horlbeck, B. Adamson, J. E. Villalta, Y. Chen, E. H. Whitehead, C. Guimaraes, B. Panning, H. L. Ploegh, M. C. Bassik, L. S. Qi, M. Kampmann and J. S. Weissman (2014). "Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation." Cell**159**(3): 647-661.

Gilbert, L. A., M. H. Larson, L. Morsut, Z. Liu, G. A. Brar, S. E. Torres, N. Stern-Ginossar, O. Brandman, E. H. Whitehead, J. A. Doudna, W. A. Lim, J. S. Weissman and L. S. Qi (2013). "CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes." Cell**154**(2): 442-451.

Goldberg, I. J. (1996). "Lipoprotein lipase and lipolysis: central roles in lipoprotein metabolism and atherogenesis." J Lipid Res**37**(4): 693-707.

Goldstein, M. and M. B. Kastan (2015). "The DNA damage response: implications for tumor responses to radiation and chemotherapy." Annu Rev Med**66**: 129-143.

Gratz, S. J., A. M. Cummings, J. N. Nguyen, D. C. Hamm, L. K. Donohue, M. M. Harrison, J. Wildonger and K. M. O'Connor-Giles (2013). "Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease." Genetics**194**(4): 1029-1035.

Gray, J. T. and S. Zolotukhin (2011). "Design and construction of functional AAV vectors." Methods Mol Biol**807**: 25-46.

Handel, E. M., K. Gellhaus, K. Khan, C. Bednarski, T. I. Cornu, F. Muller-Lerch, R. M. Kotin, R. Heilbronn and T. Cathomen (2012). "Versatile and efficient genome editing in human cells by combining zinc-finger nucleases with adeno-associated viral vectors." Hum Gene Ther**23**(3): 321-329.

Helleday, T., E. Petermann, C. Lundin, B. Hodgson and R. A. Sharma (2008). "DNA repair pathways as targets for cancer therapy." Nat Rev Cancer**8**(3): 193-204.

Hendel, A., E. J. Fine, G. Bao and M. H. Porteus (2015). "Quantifying on- and off-target genome editing." Trends Biotechnol**33**(2): 132-140.

Hilton, I. B., A. M. D'Ippolito, C. M. Vockley, P. I. Thakore, G. E. Crawford, T. E. Reddy and C. A. Gersbach (2015). "Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers." Nat Biotechnol**33**(5): 510-517.

Hirsch, M. L., L. Green, M. H. Porteus and R. J. Samulski (2010). "Self-complementary AAV mediates gene targeting and enhances endonuclease delivery for double-strand break repair." Gene Ther**17**(9): 1175-1180.

Hoeijmakers, J. H. (2001). "DNA repair mechanisms." Maturitas**38**(1): 17-22; discussion 22-13.

Hoeijmakers, J. H. (2001). "Genome maintenance mechanisms for preventing cancer." Nature**411**(6835): 366-374.

Horvath, P. and R. Barrangou (2010). "CRISPR/Cas, the immune system of bacteria and archaea." Science**327**(5962): 167-170.

Hsu, P. D., E. S. Lander and F. Zhang (2014). "Development and applications of CRISPR-Cas9 for genome engineering." Cell**157**(6): 1262-1278.

Jiang, W., H. Zhou, H. Bi, M. Fromm, B. Yang and D. P. Weeks (2013). "Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice." Nucleic Acids Res**41**(20): e188.

Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna and E. Charpentier (2012). "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Science**337**(6096): 816-821.

Kang, X., W. He, Y. Huang, Q. Yu, Y. Chen, X. Gao, X. Sun and Y. Fan (2016). "Introducing precise genetic modifications into human 3PN embryos by CRISPR/Cas-mediated genome editing." J Assist Reprod Genet**33**(5): 581-588.

Kass, E. M. and M. Jasin (2010). "Collaboration and competition between DNA double-strand break repair pathways." FEBS Lett**584**(17): 3703-3708.

Kaufmann, K. B., H. Buning, A. Galy, A. Schambach and M. Grez (2013). "Gene therapy on the move." EMBO Mol Med**5**(11): 1642-1661.

Khanna, K. K., M. F. Lavin, S. P. Jackson and T. D. Mulhern (2001). "ATM, a central controller of cellular responses to DNA damage." Cell Death Differ**8**(11): 1052-1065.

Kim, C. A. and J. M. Berg (1996). "A 2.2 A resolution crystal structure of a designed zinc finger protein bound to DNA." Nat Struct Biol**3**(11): 940-945.

Kim, Y. G., J. Cha and S. Chandrasegaran (1996). "Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain." Proc Natl Acad Sci U S A**93**(3): 1156-1160.

Kotterman, M. A. and D. V. Schaffer (2014). "Engineering adeno-associated viruses for clinical gene therapy." Nat Rev Genet**15**(7): 445-451.

Larson, M. H., L. A. Gilbert, X. Wang, W. A. Lim, J. S. Weissman and L. S. Qi (2013). "CRISPR interference (CRISPRi) for sequence-specific control of gene expression." Nat Protoc**8**(11): 2180-2196.

Lheriteau, E., A. M. Davidoff and A. C. Nathwani (2015). "Haemophilia gene therapy: Progress and challenges." Blood Rev**29**(5): 321-328.

Li, D., Z. Qiu, Y. Shao, Y. Chen, Y. Guan, M. Liu, Y. Li, N. Gao, L. Wang, X. Lu, Y. Zhao and M. Liu (2013). "Heritable gene targeting in the mouse and rat using a CRISPR-Cas system." Nat Biotechnol**31**(8): 681-683.

Li, K., G. Wang, T. Andersen, P. Zhou and W. T. Pu (2014). "Optimization of genome engineering approaches with the CRISPR/Cas9 system." PLoS One**9**(8): e105779.

Li, W., F. Teng, T. Li and Q. Zhou (2013). "Simultaneous generation and germline transmission of multiple gene mutations in rat using CRISPR-Cas systems." Nat Biotechnol**31**(8): 684-686.

Lisowski, L., S. S. Tay and I. E. Alexander (2015). "Adeno-associated virus serotypes for gene therapeutics." Curr Opin Pharmacol**24**: 59-67.

Long, C., J. R. McAnally, J. M. Shelton, A. A. Mireault, R. Bassel-Duby and E. N. Olson (2014). "Prevention of muscular dystrophy in mice by CRISPR/Cas9-mediated editing of germline DNA." Science**345**(6201): 1184-1188.

Low, B. E., M. P. Krebs, J. K. Joung, S. Q. Tsai, P. M. Nishina and M. V. Wiles (2014). "Correction of the Crb1rd8 allele and retinal phenotype in C57BL/6N mice via TALEN-mediated homology-directed repair." Invest Ophthalmol Vis Sci**55**(1): 387-395.

Luo, J., Y. Luo, J. Sun, Y. Zhou, Y. Zhang and X. Yang (2015). "Adeno-associated virus-mediated cancer gene therapy: current status." Cancer Lett**356**(2 Pt B): 347-356.

Ma, H., L. C. Tu, A. Naseri, M. Huisman, S. Zhang, D. Grunwald and T. Pederson (2016). "Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow." Nat Biotechnol**34**(5): 528-530.

Maeder, M. L., S. Thibodeau-Beganny, A. Osiak, D. A. Wright, R. M. Anthony, M. Eichtinger, T. Jiang, J. E. Foley, R. J. Winfrey, J. A. Townsend, E. Unger-Wallace, J. D. Sander, F. Muller-Lerch, F. Fu, J. Pearlberg, C. Gobel, J. P. Dassie, S. M. Pruett-Miller, M. H. Porteus, D. C. Sgroi, A. J. Iafrate, D. Dobbs, P. B. McCray, Jr., T. Cathomen, D. F. Voytas and J. K. Joung (2008). "Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification." Mol Cell**31**(2): 294-301.

Maguire, A. M., F. Simonelli, E. A. Pierce, E. N. Pugh, Jr., F. Mingozzi, J. Bennicelli, S. Banfi, K. A. Marshall, F. Testa, E. M. Surace, S. Rossi, A. Lyubarsky, V. R. Arruda, B. Konkle, E. Stone, J. Sun, J. Jacobs, L. Dell'Osso, R. Hertle, J. X. Ma, T. M. Redmond, X. Zhu, B. Hauck, O. Zelenaia, K. S. Shindler, M. G. Maguire, J. F. Wright, N. J. Volpe, J. W. McDonnell, A. Auricchio, K. A. High and J. Bennett (2008). "Safety and efficacy of gene transfer for Leber's congenital amaurosis." N Engl J Med**358**(21): 2240-2248.

Mak, A. N., P. Bradley, R. A. Cernadas, A. J. Bogdanove and B. L. Stoddard (2012). "The crystal structure of TAL effector PthXo1 bound to its DNA target." Science**335**(6069): 716-719.

Makarova, K. S., Y. I. Wolf, O. S. Alkhnbashi, F. Costa, S. A. Shah, S. J. Saunders, R. Barrangou, S. J. Brouns, E. Charpentier, D. H. Haft, P. Horvath, S. Moineau, F. J. Mojica, R. M.

Terns, M. P. Terns, M. F. White, A. F. Yakunin, R. A. Garrett, J. van der Oost, R. Backofen and E. V. Koonin (2015). "An updated evolutionary classification of CRISPR-Cas systems." Nat Rev Microbiol**13**(11): 722-736.

Maruyama, T., S. K. Dougan, M. C. Truttmann, A. M. Bilate, J. R. Ingram and H. L. Ploegh (2015). "Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining." Nat Biotechnol**33**(5): 538-542.

Matos, J. and S. C. West (2014). "Holliday junction resolution: regulation in space and time." DNA Repair (Amst)**19**: 176-181.

Matrai, J., M. K. Chuah and T. VandenDriessche (2010). "Recent advances in lentiviral vector development and applications." Mol Ther**18**(3): 477-490.

McVey, M. and S. E. Lee (2008). "MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings." Trends Genet**24**(11): 529-538.

Miller, J. C., S. Tan, G. Qiao, K. A. Barlow, J. Wang, D. F. Xia, X. Meng, D. E. Paschon, E. Leung, S. J. Hinkley, G. P. Dulay, K. L. Hua, I. Ankoudinova, G. J. Cost, F. D. Urnov, H. S. Zhang, M. C. Holmes, L. Zhang, P. D. Gregory and E. J. Rebar (2011). "A TALE nuclease architecture for efficient genome editing." Nat Biotechnol**29**(2): 143-148.

Miller, N. (2012). "Glybera and the future of gene therapy in the European Union." Nature Reviews Drug Discovery**11**(5): 419-419.

Mimitou, E. P. and L. S. Symington (2008). "Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing." Nature**455**(7214): 770-774.

Moehle, E. A., J. M. Rock, Y. L. Lee, Y. Jouvenot, R. C. DeKelver, P. D. Gregory, F. D. Urnov and M. C. Holmes (2007). "Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases." Proc Natl Acad Sci U S A**104**(9): 3055-3060.

Mullaney, J. M., R. E. Mills, W. S. Pittard and S. E. Devine (2010). "Small insertions and deletions (INDELs) in human genomes." Hum Mol Genet**19**(R2): R131-136.

Niu, Y., B. Shen, Y. Cui, Y. Chen, J. Wang, L. Wang, Y. Kang, X. Zhao, W. Si, W. Li, A. P. Xiang, J. Zhou, X. Guo, Y. Bi, C. Si, B. Hu, G. Dong, H. Wang, Z. Zhou, T. Li, T. Tan, X. Pu, F. Wang, S. Ji, Q. Zhou, X. Huang, W. Ji and J. Sha (2014). "Generation of gene-modified cynomolgus monkey via Cas9/RNA-mediated gene targeting in one-cell embryos." Cell**156**(4): 836-843.

Oliveira, N. G., M. Castro, A. S. Rodrigues, O. M. Gil, J. M. Toscano-Rico and J. Rueff (2002). "DNA-PK inhibitor wortmannin enhances DNA damage induced by bleomycin in V79 Chinese hamster cells." Teratog Carcinog Mutagen**22**(5): 343-351.

Pardo, B., B. Gomez-Gonzalez and A. Aguilera (2009). "DNA repair in mammalian cells: DNA double-strand break repair: how to fix a broken relationship." Cell Mol Life Sci**66**(6): 1039-1056.

Pattanayak, V., C. L. Ramirez, J. K. Joung and D. R. Liu (2011). "Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection." Nat Methods**8**(9): 765-770.

Pavletich, N. P. and C. O. Pabo (1991). "Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A." Science**252**(5007): 809-817.

Pearson, S., H. Jia and K. Kandachi (2004). "China approves first gene therapy." Nat Biotechnol**22**(1): 3-4.

Peng, Z. (2005). "Current status of gendicine in China: recombinant human Ad-p53 agent for treatment of cancers." Hum Gene Ther**16**(9): 1016-1027.

Pham, H., N. A. Kearns and R. Maehr (2016). "Transcriptional Regulation with CRISPR/Cas9 Effectors in Mammalian Cells." Methods Mol Biol**1358**: 43-57.

Pierce, A. J., R. D. Johnson, L. H. Thompson and M. Jasin (1999). "XRCC3 promotes homology-directed repair of DNA damage in mammalian cells." Genes Dev**13**(20): 2633-2638.

Pingoud, A., G. G. Wilson and W. Wende (2014). "Type II restriction endonucleases--a historical perspective and more." Nucleic Acids Res**42**(12): 7489-7527.

Pourzand, C. and P. Cerutti (1993). "Genotypic mutation analysis by RFLP/PCR." Mutat Res**288**(1): 113-121.

Qiu, P., H. Shandilya, J. M. D'Alessio, K. O'Connor, J. Durocher and G. F. Gerard (2004). "Mutation detection using Surveyor nuclease." Biotechniques**36**(4): 702-707.

Ramamoorth, M. and A. Narvekar (2015). "Non viral vectors in gene therapy- an overview." J Clin Diagn Res**9**(1): GE01-06.

Ran, F. A., P. D. Hsu, J. Wright, V. Agarwala, D. A. Scott and F. Zhang (2013). "Genome engineering using the CRISPR-Cas9 system." Nat Protoc**8**(11): 2281-2308.

Renkawitz, J., C. A. Lademann and S. Jentsch (2014). "Mechanisms and principles of homology search during recombination." Nat Rev Mol Cell Biol**15**(6): 369-383.

Richardson, C. D., G. J. Ray, M. A. DeWitt, G. L. Curie and J. E. Corn (2016). "Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA." Nat Biotechnol**34**(3): 339-344.

Rodgers, K. and M. McVey (2016). "Error-Prone Repair of DNA Double-Strand Breaks." J Cell Physiol**231**(1): 15-24.

Rosenzweig, K. E., M. B. Youmell, S. T. Palayoor and B. D. Price (1997). "Radiosensitization of human tumor cells by the phosphatidylinositol3-kinase inhibitors wortmannin and LY294002 correlates with inhibition of DNA-dependent protein kinase and prolonged G2-M delay." Clin Cancer Res**3**(7): 1149-1156.

Sancar, A., L. A. Lindsey-Boltz, K. Unsal-Kacmaz and S. Linn (2004). "Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints." Annu Rev Biochem**73**: 39-85.

Sander, J. D., E. J. Dahlborg, M. J. Goodwin, L. Cade, F. Zhang, D. Cifuentes, S. J. Curtin, J. S. Blackburn, S. Thibodeau-Beganny, Y. Qi, C. J. Pierick, E. Hoffman, M. L. Maeder, C. Khayter, D. Reyon, D. Dobbs, D. M. Langenau, R. M. Stupar, A. J. Giraldez, D. F. Voytas, R. T. Peterson, J. R. Yeh and J. K. Joung (2011). "Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA)." Nat Methods**8**(1): 67-69.

Santiago-Ortiz, J. L. and D. V. Schaffer (2016). "Adeno-associated virus (AAV) vectors in cancer gene therapy." J Control Release.

Sfeir, A. and L. S. Symington (2015). "Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway?" Trends Biochem Sci**40**(11): 701-714.

Smith, J., M. Bibikova, F. G. Whitby, A. R. Reddy, S. Chandrasegaran and D. Carroll (2000). "Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains." Nucleic Acids Res**28**(17): 3361-3369.

Srivastava, M., M. Nambiar, S. Sharma, S. S. Karki, G. Goldsmith, M. Hegde, S. Kumar, M. Pandey, R. K. Singh, P. Ray, R. Natarajan, M. Kelkar, A. De, B. Choudhary and S. C. Raghavan (2012). "An inhibitor of nonhomologous end-joining abrogates double-strand break repair and impedes cancer progression." Cell**151**(7): 1474-1487.

Srivastava, M. and S. C. Raghavan (2015). "DNA double-strand break repair inhibitors as cancer therapeutics." Chem Biol**22**(1): 17-29.

Stieger, K. and B. Lorenz (2008). "[The treatment of inherited dystrophies and neovascular disorders of the retina by rAAV-mediated gene therapy]." Klin Monbl Augenheilkd**225**(12): 1009-1023.

Stieger, K. and B. Lorenz (2014). "[Specific gene therapy for hereditary retinal dystrophies - an update]." Klin Monbl Augenheilkd**231**(3): 210-215.

Stoessl, A. J. (2014). "Gene therapy for Parkinson's disease: a step closer?" Lancet**383**(9923): 1107-1109.

Sturzenegger, A., K. Burdova, R. Kanagaraj, M. Levikova, C. Pinto, P. Cejka and P. Janscak (2014). "DNA2 cooperates with the WRN and BLM RecQ helicases to mediate long-range DNA end resection in human cells." J Biol Chem**289**(39): 27314-27326.

Tang, H., K. L. Kuhen and F. Wong-Staal (1999). "Lentivirus replication and regulation." Annu Rev Genet**33**: 133-170.

Tiscornia, G., O. Singer and I. M. Verma (2006). "Production and purification of lentiviral vectors." Nat Protoc**1**(1): 241-245.

Tomimatsu, N., B. Mukherjee, M. Catherine Hardebeck, M. Ilcheva, C. Vanessa Camacho, J. Louise Harris, M. Porteus, B. Llorente, K. K. Khanna and S. Burma (2014). "Phosphorylation of EXO1 by CDKs 1 and 2 regulates DNA end resection and repair pathway choice." Nat Commun**5**: 3561.

Truong, L. N., Y. Li, E. Sun, K. Ang, P. Y. Hwang and X. Wu (2014). "Homologous recombination is a primary pathway to repair DNA double-strand breaks generated during DNA rereplication." J Biol Chem**289**(42): 28910-28923.

Urnov, F. D., J. C. Miller, Y. L. Lee, C. M. Beausejour, J. M. Rock, S. Augustus, A. C. Jamieson, M. H. Porteus, P. D. Gregory and M. C. Holmes (2005). "Highly efficient endogenous human gene correction using designed zinc-finger nucleases." Nature**435**(7042): 646-651.

Urnov, F. D., E. J. Rebar, M. C. Holmes, H. S. Zhang and P. D. Gregory (2010). "Genome editing with engineered zinc finger nucleases." Nat Rev Genet**11**(9): 636-646.

van der Oost, J., E. R. Westra, R. N. Jackson and B. Wiedenheft (2014). "Unravelling the structural and mechanistic basis of CRISPR-Cas systems." Nat Rev Microbiol**12**(7): 479-492.

Vriend, L. E., M. Jasin and P. M. Krawczyk (2014). "Assaying break and nick-induced homologous recombination in mammalian cells using the DR-GFP reporter and Cas9 nucleases." Methods Enzymol**546**: 175-191.

Wah, D. A., J. Bitinaite, I. Schildkraut and A. K. Aggarwal (1998). "Structure of FokI has implications for DNA cleavage." Proc Natl Acad Sci U S A**95**(18): 10564-10569.

Wang, H., H. Yang, C. S. Shivalila, M. M. Dawlaty, A. W. Cheng, F. Zhang and R. Jaenisch (2013). "One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering." Cell**153**(4): 910-918.

Wei, C., J. Liu, Z. Yu, B. Zhang, G. Gao and R. Jiao (2013). "TALEN or Cas9 - rapid, efficient and specific choices for genome modifications." J Genet Genomics**40**(6): 281-289.

Willmore, E., S. de Caux, N. J. Sunter, M. J. Tilby, G. H. Jackson, C. A. Austin and B. W. Durkacz (2004). "A novel DNA-dependent protein kinase inhibitor, NU7026, potentiates the cytotoxicity of topoisomerase II poisons used in the treatment of leukemia." Blood**103**(12): 4659-4665.

Wilson, J. M. (2005). "Gendicine: the first commercial gene therapy product." Hum Gene Ther**16**(9): 1014-1015.

Wright, A. V., J. K. Nunez and J. A. Doudna (2016). "Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering." Cell**164**(1-2): 29-44.

Wu, J., K. Kandavelou and S. Chandrasegaran (2007). "Custom-designed zinc finger nucleases: what is next?" Cell Mol Life Sci**64**(22): 2933-2944.

Wu, S., G. Ying, Q. Wu and M. R. Capecchi (2008). "A protocol for constructing gene targeting vectors: generating knockout mice for the cadherin family and beyond." Nat Protoc**3**(6): 1056-1076.

Yang, H., H. Wang, C. S. Shivalila, A. W. Cheng, L. Shi and R. Jaenisch (2013). "One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering." Cell**154**(6): 1370-1379.

Yanik, M., B. Muller, F. Song, J. Gall, F. Wagner, W. Wende, B. Lorenz and K. Stieger (2016). "In vivo genome editing as a potential treatment strategy for inherited retinal dystrophies." Prog Retin Eye Res.

Yun, M. H. and K. Hiom (2009). "CtIP-BRCA1 modulates the choice of DNA double-strand-break repair pathway throughout the cell cycle." Nature**459**(7245): 460-463.

Zhu, Z., W. H. Chung, E. Y. Shim, S. E. Lee and G. Ira (2008). "Sgs1 helicase and two nucleases Dna2 and Exo1 resect DNA double-strand break ends." Cell**134**(6): 981-994.

Acknowledgments

An dieser Stelle möchte ich mich recht herzlich bei allen Personen bedanken, die mich bei der Erstellung dieser Arbeit unterstützt haben.

Herrn Prof. Dr. Peter Friedhoff danke ich für die Übernahme meiner Betreuung als mein Doktorvater, seine wissenschaftliche Unterstützung und die Anregungen zu meiner Doktorarbeit.

Bei Herrn Prof. Dr. Dr. Knut Stieger bedanke ich mich für die Bereitstellung des Themas und die gute Betreuung meiner Doktorarbeit. Dabei möchte ich mich insbesondere für die Unterstützung, die guten Ratschläge und die schnelle Durchsicht meiner Arbeit bedanken.

Bei Frau Prof. Dr. Birgit Lorenz bedanke ich mich für die Bereitstellung des Arbeitsplatzes und die wissenschaftliche Unterstützung während der gesamten Zeit.

Ganz besonders möchte ich mich bei Herrn PD Dr. Markus Preising für die wissenschaftliche Unterstützung bedanken. Ich danke Dir für die Einarbeitung in die Laborarbeit, die Begleitung meiner gesamten Promotionszeit und die zahlreichen Ratschläge und Tipps.

An dieser Stelle gilt mein Dank ebenfalls Frau Dr. Brigitte Müller, Frau Dr. Claudia Lopez und Herrn Dr. Mert Yanik, für die stets gute und hilfsbereite Atmosphäre, für die wissenschaftlichen Tipps und Diskussionen.

Außerdem möchte ich mich bei Frau Dr. Nelli Baal und Frau Gabriela Michel, FACS Core Facility Klinische Immunologie, für die Bereitstellung und intensive Einarbeitung am FACS bedanken.

Ein ganz besonderes und herzliches Dankeschön an Franzi und Jacky, die mich während der Promotionszeit begleitet haben. Ich danke Euch für die Unterstützung, die Motivation und die tolle Zeit.

Ich danke Tobi und Bella für die Einarbeitung in die Laborarbeit, für die jederzeitige Ansprechbarkeit und Hilfestellung während meiner Master- und Promotionszeit.

Des Weiteren möchte ich mich bei allen Mitarbeitern des Labors für molekulare Ophthalmologie der Augenklinik für die gute Zusammenarbeit, die Hilfsbereitschaft und das angenehme Arbeitsklima bedanken.

Abschließend möchte ich mich ganz besonders bei meinen Eltern und meiner Familie für die Unterstützung und Motivation während der gesamten Zeit bedanken und dafür, dass Ihr mir all das ermöglicht habt.

最后，我要由衷的感谢父母，家人和朋友的支持。