Charles University

Faculty of Science

Study programme: Molecular Biology and Biochemistry of Organisms

Pavlína Faltínová

CRISPR/Cas genome editing in human disease models and the translation into therapies

Editace genomu pomocí CRISPR/Cas na modelech lidských onemocnění a jejich využití v terapii

Bachelor's thesis

Supervisor: Mgr. Michaela Krupková, Ph.D.

Prague, 2024

Acknowledgements

First, I would like to express my thanks to my supervisor Mgr. Michaela Krupková, Ph.D. for her patience, valuable advice and all the help she provided me with while I was working on this thesis. My gratitude also belongs to my family and friends who emotionally supported me throughout my studies and, especially, throughout the process of writing my thesis.

Prohlášení

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze, 27. 4. 2024

Pavlína Faltínová

Abstract

Rare diseases, as their name indicates, individually affect only a low number of people around the world. Due to their low prevalence, finding appropriate therapy is very difficult. Insufficient understanding of the molecular causalities and mechanisms accompanying these disorders and the inability to conduct clinical studies to the usual extent due to the low occurrence of rare diseases belong to the main problems hindering the development of proper treatment. The creation of mouse models is a promising way to solve these difficulties since mice have numerous qualities necessary for modelling human diseases. CRISPR/Cas9 enables scientists to make precise changes in the genome by employing Cas9 nuclease which creates double-strand breaks in the DNA after the specifically designed guide RNA leads it to the site of interest. If the sequence of the chosen site is known, the possible edits can be installed almost anywhere in the genome and, moreover, their repertoire is practically endless. The use of the CRISPR/Cas9 technology proved to be perfect for creating mouse models of rare diseases as most of these disorders are caused by genetic mutations that this method is fully capable of mimicking. This thesis focuses on strategies used in creating such mouse models with the CRISPR/Cas9 system and summarizes their detailed mechanisms.

Keywords: rare diseases, mouse model, CRISPR/Cas9, knock-out, conditional knock-out, knock-in, transgenic mice, base editing, prime editing

Abstrakt

Vzácná onemocnění, jak už jejich název napovídá, individuálně postihují jen malý počet lidí na celém světě. Vzhledem k jejich nízké prevalenci je nalezení vhodné terapie velmi obtížné. Neuspokojivé pochopení molekulárních příčin a mechanismů, které je provázejí, a nemožnost provádět klinické studie v obvyklém rozsahu v důsledku nízkého výskytu vzácných onemocnění patří k hlavním problémům, které zbržďují vývoj vhodné léčby. Tvorba myších modelů je slibnou cestou k vyřešení těchto obtíží, neboť myši mají řadu vlastností nezbytných pro modelování lidských onemocnění. CRISPR/Cas9 umožňuje vědcům provádět přesné změny v genomu pomocí nukleázy Cas9, která vytváří dvouřetězcové zlomy v DNA poté, co ji speciálně navržená guide RNA přivede k místu zájmu. Pokud je sekvence zvoleného místa známá, lze možné úpravy zavést téměř kamkoliv do genomu a jejich repertoár je navíc prakticky nekonečný. Využití CRISPR/Cas9 technologie se ukázalo být ideální pro tvorbu myších modelů vzácných onemocnění, neboť většina těchto poruch je způsobena genetickými mutacemi, které dokáže tato metoda plně napodobit. Tato práce se zaměřuje na strategie využívané při tvorbě myších modelů pomocí systému CRISPR/Cas9 a shrnuje jejich podrobné mechanismy.

Klíčová slova: vzácná onemocnění, myší model, CRISPR/Cas9, knock-out, kondicionální knock-out, knock-in, transgenní myši, editace bází, primární editace

List of abbreviations

A…… adenosine-5’-phosphate nucleotide

AAG…….. alkyl adenine DNA glycosylase

ABE……... adenine base editor

ACBE…… adenine and cytidine base editor

AID…… activation-induced cytidine deaminase

AP lyase…… apurinic or apyrimidinic site lyase

APOBEC……... apolipoprotein B mRNA editing enzyme, catalytic polypeptide

BE-PLUS……...base editor for programming larger C to U (T) scope

BER …….. base excision repair

C…… cytidine-5’-phosphate nucleotide

Cas9……... CRISPR-associated protein 9

Cas9n……Cas9 nickase

CBE…..…….. cytidine base editor

CGBE …….. C-to-G base editor

Cre…… cyclization recombinase

CRISPR……. Clustered Regularly Interspaced Short Palindromic Repeats

CRISPR-HITI…… CRISPR/Cas9 homology-independent targeted insertion

CRISPR-PITCh……. CRISPR/Cas9-mediated Precise Integration into Target Chromosome

crRNA……... CRISPR RNA

dCas9……dead Cas9

DNA…….. deoxyribonucleic acid

DSB……... double-strand break

dsDNA……...double-stranded DNA

Easi-CRISPR……Efficient additions with ssDNA inserts-CRISPR

epegRNA……... engineered pegRNA

Flp……. flippase recombinase

FRT……FLP recombination target

G…… guanosine-5’-phosphate nucleotide

gRNA…….guide RNA

HDR…….. homology-directed repair

HOPE…….homologous 3′ extension mediated prime editor

I……..inosine-5’-phosphate nucleotide

IKCM…….International Knockout Mouse Consortium

IPCM……. International Phenotyping Mouse Consortium

ivTRT…….*in vitro* transcription and reverse transcription

Kibl……Knock-in blunt ligation

loxP……locus of crossing-over, P1

MMEJ……microhomology-mediated end joining

M-MLV……..Moloney murine leukaemia virus

mRNA…….. messenger RNA

NHEJ……. non-homologous end joining

NLS…….... nuclear localization signal

NMD……. nonsense-mediated decay

PAM…….. protospacer-adjacent motif

PBS……primer binding site

PCR……... polymerase chain reaction

PE…….. prime editor

PEDAR……..PE-Cas9-based deletion and repair

pegRNA…….prime editing guide RNA

PmCDA1……... *Petromyzon marinus* cytidine deaminase 1

poly(A)tail……polyadenylic acid tail

RNA……...ribonucleic acid

SCON..…….. Short Conditional intrON

sgRNA……... single-guide RNA

*Sp*Cas9……... *Streptococcus pyogenes* Cas9

SSA……single-strand annealing

ssDNA …….. single-stranded DNA

ssODN …….. single-stranded oligonucleotide

T…….thymidine-5’-phosphate nucleotide

TadA……..tRNA adenosine deaminase A

TALEN……..Transcription Activator-Like Effector Nuclease

Tild-CRISPR……. Targeted integration with linearized dsDNA-CRISPR

TKIT……..Targeted Knock-In with Two guides

tracrRNA……... transactivating CRISPR RNA

tRNA……. transfer RNA

U……uridine-5’-phosphate nucleotide

UGI……uracil DNA glycosylase inhibitor

UNG…….. uracil N-glycosylase

UTR……... untranslated region

ZFN……... Zinc-Finger Nuclease

Outline

1. Introduction 1
2. Rare diseases 2
3. Mouse as a model organism for rare diseases 3
4. CRISPR/Cas9 5
4.1. Components and functioning 5
4.2. Modification of the Cas9 nuclease domains 7
5. Genome editing methods and strategies 7
5.1. Gene knock-out 7
5.1.1 Single gRNA strategy 7
5.1.1.1. Possible challenges 8
5.1.2. Dual gRNA strategies 9
5.1.3. Double-nicking strategy 10
5.2. Gene knock-in 11
5.2.1. Homology-dependent gene knock-in 11
5.2.2. Homology-independent gene knock-in 14
5.2.3. Transgenic mice 16
5.2.4. Gene knock-out mediated by homology-directed repair 18
5.2.5. Conditional knock-out 18
5.2.5.1. Creation of conditional allele 19
5.3. Base editing 22
5.3.1. Base editors 23
5.3.1.1. Cytidine base editors 23
5.3.1.2. Adenine base editors 25
5.3.1.3. C to G base editor 1 26
5.3.1.4. Adenine and cytidine base editor 27
5.3.2. Editing window 27
5.4. Prime editing 28
5.4.1. Prime editors 29
5.4.2. Alteration of pegRNA 30
5.4.3. Prime editing using two pegRNAs 31
6. Conclusion 32
7. List of literature 34

1. Introduction

Rare diseases are a group of extremely heterogeneous disorders mostly of genetic origin individually affecting only a handful of people worldwide (EU Regulation on Orphan Medicinal Products, 1999, Nguengang Wakap *et al*., 2019, Rare Disease Act, 2002). With only a few affected individuals it is rather difficult to conduct any research to accurately describe the causality and clinical manifestation of these diseases resulting in a major lack of information in this field. Furthermore, the unique nature of rare diseases leads most pharmaceutical companies to lose interest in developing drugs for their treatment. Important steps to address these difficulties have already been taken (Orphan Drug Act, 1983). However, there is still a long way to go to make essential information accessible to all individuals suffering from rare diseases.

Animal models have been widely utilized to help researchers properly understand the causalities of rare diseases which is a necessity for the effective development of appropriate therapeutical strategies. Additionally, modelling on animals is also beneficial since clinical trials required for evaluating new drugs cannot be effectively performed on people as the number of patients is very limited (EU Regulation on Orphan Medicinal Products, 1999, Orphan Drug Act, 1983). Mice were proved to be the most convenient organisms for modelling human diseases for numerous reasons, such as their low cost, high speed of reproduction, ease of manipulation and, importantly, relative translatability of the results acquired during research to humans. Nevertheless, creating appropriate models corresponding to specific rare diseases is not an easy task and requires advanced scientific approaches.

CRISPR/Cas9 is a relatively novel gene editing technology which has, however, already gained extreme popularity among researchers for its relatively low cost, simplicity and, finally, high efficiency. The original defence mechanism used by prokaryotes against invading pathogens was first transformed and utilized as a technique for genome alteration by Jennifer Doudna, Emmanuelle Charpentier and their team (Jinek *et al*., 2012). This strategy employs the Cas9 nuclease to create double-strand breaks (DSBs) in the genome which are, subsequently, fixed by cellular repair mechanisms with various results (Fu *et al*, 2021, Jinek *et al*., 2012). The CRISPR/Cas9 technology is able to create many different edits such as frameshift mutations and deletions resulting in gene knock-out, insertions of genetic sequences with varying lengths, single and multiple base substitutions, etc (Anzalone *et al*., 2019, Canver *et al*., 2014, Komor *et al*., 2016, Raveux *et al*., 2017, Yen *et al*., 2014). This approach has been successfully utilized for the creation of mouse models of rare diseases on numerous occasions enabling scientists to develop a treatment for such disorders more effectively (Borrás *et al*., 2020, Duan *et al*., 2016, Liang *et al*., 2017, Liang *et al*., 2018, Lin J. *et al*., 2021, Nakagawa *et al*., 2016, Nutter *et al*., 2019, Qian *et al*., 2023, Rauch *et al*., 2018, Syding *et al*., 2022, Zhang Y. *et al*., 2022).

2. Rare diseases

Rare diseases, sometimes also termed orphan diseases, are medical conditions individually affecting only a very small part of the world population. It is rather difficult to describe rare diseases accurately as there is no universally accepted definition. Depending on the area or the context, in which they are used, numerous different definitions of such conditions exist. Most of those at least partly consider the disease point prevalence which proved to be a globally preferred attribute, however, many use various qualitative norms as a main describing trait (Richter *et al*., 2015). In the United States, rare diseases are viewed as conditions affecting less than 200,000 people according to the Rare disease Act of 2002 (Rare disease Act, 2002). Nevertheless, the point prevalence should be no more than 5 per 10,000 people for a disease to be classified as rare in Europe (EU Regulation on Orphan Medicinal Products, 2000). This definition based on point prevalence showed to be more convenient as it considers population growth in time. The US definition based on the absolute number of people affected leads to a decrease in point prevalence over time as the population grows, which can cause further difficulties (Nguengang Wakap *et al*., 2020).

To keep up with the increasing number of newly discovered rare diseases, Orphanet was established in 1997 by the French National Institute of Health and Medical Research. Orphanet is a database containing information on all registered rare diseases and its main purpose is to store valuable data and to make them easily accessible to healthcare specialists and the general public. This portal is updated every month in order to include new rare diseases or to further specify the classification of current disorders as new details are discovered. Orphanet uses data from various sources such as scientific publications or workshops with experts in the field. The cooperation among research groups as well as among whole countries is critical for the database and enables it to include a large amount of reliable information. Therefore, Orphanet has been funded by the European Commission since 2000, and, currently, almost forty different countries contribute to the data collection (Rath *et al*., 2012).

Although the exact number of rare diseases differs tremendously among various sources, Orphanet includes descriptions of more than 6100 unique rare diseases. Moreover, almost 72 % of these conditions are considered to be of genetic origin and almost 70 % are paediatric onset. Nguengang Wakap *et al*. conducted research in an effort to estimate the global cumulative point prevalence of these disorders. They considered the rare diseases registered in the Orphanet database excluding rare cancers, infectious diseases and poisonings as well as rare diseases with unknown point prevalence. Using data from the remaining 67.6 % of diseases, they calculated the global point prevalence as 3.5-5.9 %. This represents approximately 262.9 to 446.2 million people suffering from rare diseases worldwide, considering the size of the population in 2017 when the research was conducted. Furthermore, as the study knowingly excluded some of the registered rare diseases, the real number is likely to be even higher. It is, therefore, important to note that while each rare disease

individually affects only a small number of people, collectively they pose a major challenge to modern medicine (Nguengang Wakap *et al*., 2020).

Rare diseases contain large amounts of extremely heterogeneous disorders with various molecular aetiology resulting in different clinical manifestations. As mentioned above, more than 70 % of these conditions are of genetic origin. Nevertheless, even with this specification the causes of individual disorders differ immensely including for example base substitutions, insertions or deletions, translocations, etc. Moreover, a single disease can be caused by multiple molecular mechanisms making it very complex and heterogenous on its own. On the other hand, many disorders often share the same causality even though their effects on a patient's phenotype differ. Thus, it would be beneficial to group patients according to the disease aetiology rather than the manifestations for clinical trials. That is because a single drug can be used to target the molecular causality shared by multiple disorders and the same therapeutical approach, therefore, has the potential to treat more than one disease simultaneously (Brooks *et al*., 2014, Rath *et al*., 2012).

There is, unfortunately, a general lack of information on rare diseases. Because of their low prevalence, not enough participants are available for conducting research which could help accurately describe the causalities and provide data for the development of new treatment (Richter *et al*., 2015). Furthermore, financial loss is expected for pharmaceutical companies involved in creating orphan drugs, which is a term used for drugs specifically designed to cure rare diseases. The reason for this is that the high cost of development would by far exceed the profit generated from sales. The main purpose of the Orphan Drug Act authorized in the United States in 1983 was to provide financial incentives for pharmaceutical companies to enhance the development of such drugs. Over the forty years, it is in effect, a large number of new drugs for rare diseases were found and between 4 to 6 % of rare diseases already have an approved drug on the market. (Fermaglich and Miller, 2023, Orphan Drug Act, 1983).

3. Mouse as a model organism for rare diseases

Accurate description of causalities and mechanisms connected to various human diseases as well as effective development of new therapeutical approaches requires profound modelling. In clinical practice, modelling is usually done either *in vitro* using cell cultures and organ cultures or *in vivo* using animal models. Nevertheless, *in vitro* experiments, although useful for obtaining preliminary information, are not as suitable for precise modelling of human diseases for various reasons. For example, when a chemical substance is transferred *in vivo* it encounters many natural barriers which do not exist in cell cultures and are generally difficult to reproduce *in vitro*. Moreover, a given chemical substance is most likely specifically metabolized inside the animal's body, however, this is not the case when using cell cultures. Finally, *in vitro* experimenting cannot mimic the overall complexity and structure of the human body nearly as much as animal models, which makes the translatability

of the results obtained from these types of studies rather low. *In vivo* research and accurate animal models are, thus, necessary for gaining critical information which could help develop effective treatments for various human disorders (Garattini and Grignaschi, 2017).

The first and most important step in animal modelling is choosing a suitable organism to use. In modern research, rodents and especially mice proved to be the most convenient organism for modelling human diseases for several reasons. First of all, mice are small and easy to breed and manipulate, as they do not require much space or resources. They are cost-effective and highly available even in large numbers which is crucial as many subjects are often needed in a single experiment for scientists to be able to statistically evaluate the results of research. Additionally, mice have a much shorter generation time and their lifespan is accelerated compared to humans, thus, researchers are able to obtain experimental results in a relatively short period of time (Vandamme, 2015). The mouse genetic background is also profoundly explored, as its whole genome has been sequenced. As a result, the designing and modelling processes are highly simplified (Mouse Genome Sequencing Consortium, 2002). Furthermore, genetic engineering is possible in mice, which is a necessary quality that allows scientists to create an enormous diversity of mouse models.

Finally, an extremely important feature is the translatability of the experimental results to humans. One of the main reasons why mice were chosen as model organisms for medical research is the similarity of their genome to the human one. Numerous genomic regions and sequences are highly conserved among all mammals and can be found in both mice and humans in an unmodified state. Moreover, the absolute majority of mouse genes have at least one homolog in the human genome. Therefore, the results obtained from experiments using mouse models can be, to some extent, applied to humans (Mouse Genome Sequencing Consortium, 2002).

In 2007, the International Knockout Mouse Consortium (IKCM) launched a project whose objective was to create knock-out mutations in every known coding gene of the mouse genome. Over the years, IKCM generated countless edited mouse embryonic stem cell clones. The majority of these mutations were performed on embryonic stem cell lines originating from C57BL/6, one of the best-characterized inbred mouse strains widely utilized in the scientific field. (Bradley *et al*., 2012, Pettitt *et al*., 2009).

The International Phenotyping Mouse Consortium (IPCM), launched in 2011, utilizes mutant embryonic stem cell lines created by IKCM to generate and, subsequently, phenotype specific knock-out mouse strains. Its main aim is to determine the exact function of mutated mouse genes and the function of their mammalian, especially human, homologs. Many of these specifically designed mouse strains are successfully used as model organisms for human diseases in medical research (Brown and Moore, 2012).

4. CRISPR/Cas9

CRISPR/Cas9 is an innovative and currently the most studied gene-editing tool in the scientific field. Its efficiency, relatively low cost and simple design make it an extremely powerful and beneficial way to edit genes. The CRISPR/Cas9 system differs from its ancestors such as zinc-finger nucleases (ZFNs) or transcription activator-like effector nucleases (TALENs) mainly in the way it targets the chosen DNA sequence. While ZFNs and TALENs use specific sequences of amino acids to bind to the target genomic locus, the CRISPR/Cas9 technology relies on short RNA containing bases complementary to the target site (Boch *et al*., 2009, Jinek *et al*., 2012, Mani *et al*., 2005). Since predicting the amino-acid sequence, which would precisely bind to the chosen locus, is rather difficult, the CRISPR/Cas9 system is viewed as a much more suitable option.

4.1. Components and functioning

CRISPR/Cas9 originates from an intricate immunity system against bacteriophages found in many bacteria and archaea (Barrangou *et al*. 2007). It obtained its name from the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) array, which is an area in the prokaryote genome essential for this antiviral system to work. (Jansen *et al*., 2002).

This system consists of CRISPR-associated protein 9 (Cas9), CRISPR RNA (crRNA), which is precisely designed to target the DNA sequence of interest, and transactivating CRISPR RNA (tracrRNA). TracrRNA binds to crRNA in a base complementary fashion and then enables the association with the Cas9 protein (Jinek *et al*., 2012).

The crRNA:tracrRNA complex guides the Cas9 nuclease to target locus in the genome through complementary base pairing between the chosen sequence and crRNA. Critical for successful binding is the recognition of the protospacer-adjacent motif (PAM) located downstream of the target site by the Cas9 protein (Mojica *et al*., 2009). Today, the most frequently used Cas9 protein is *Sp*Cas9 extracted from *Streptococcus pyogenes*. *Sp*Cas9 recognizes the PAM sequence of NGG, which occurs relatively frequently and is common in the genomes of many organisms, therefore, significantly facilitating the use of the CRISPR/Cas9 technology (Hsu *et al*., 2013). The binding and complementary base pairing of the complex to the DNA promotes the formation of an R-loop structure (Szczelkun *et al*., 2014). Two of the Cas9 domains called HNH and RuvC display nuclease activity and their function is to cleave the complementary strand to crRNA and the non-complementary strand, respectively (Jinek *et al*., 2012). The establishment of an R-loop activates these domains and catalyzes the cleavage of double-stranded DNA, thus, creating a DSB (Sternberg *et al*., 2014). Subsequently, the conserved cell repair mechanisms such as non-homologous end joining (NHEJ) or homology-directed repair (HDR) are activated (Fu *et al*., 2021).

NHEJ is a predominant repair pathway in mammalian cells because it is active in all stages of the cell cycle, although it preferentially functions in G1 and early S phases (Takata *et al*., 1998).

The blunt DNA ends are rejoined with only minor processing during the NHEJ pathway. This mechanism is, however, error-prone and often results in the formation of various mutations, especially small insertions or deletions (indels) (Fu *et al*., 2021). HDR, on the other hand, is much more precise and repairs the damaged DNA with the help of a homologous template, often a sister chromatid. Nevertheless, this template is naturally available only in the S and G2 phases of the cell cycle which restricts the pathway from being active during other stages (Takata *et al*., 1998).

Furthermore, the structure of the dual crRNA:tracrRNA complex suggested that connecting both of these RNAs into a single chimera while simultaneously preserving their properties could be possible. The 3' end of crRNA was, therefore, fused with the 5' end of tracrRNA to create single-guide RNA (sgRNA). This novel RNA consists of a recognition sequence situated at the 5' end, which is responsible for binding to the target site in the DNA, followed by a hairpin structure ensuring the proper base pairing between former crRNA and tracrRNA. Fusing two individual RNAs into single chimeric sgRNA tremendously simplified the designing process and made the CRISPR/Cas9 system even more convenient for gene editing (Jinek *et al*., 2012).

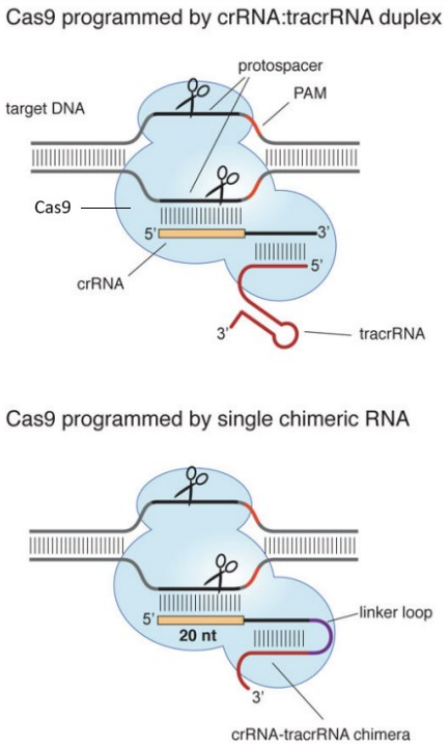

Figure 1: Schematic of Cas9-associated RNA binding to the target DNA site.
Top: Cas9 is guided to the target genomic site by two individual RNAs: crRNA, which binds to the target site in a base-complementary fashion, and tracrRNA. Bottom: By fusing crRNA and tracrRNA via a linker loop, a single-guide RNA is created. This chimera effectively leads the Cas9 protein to the target DNA site and simplifies the designing process. The scissors symbolize the cleavage sites in both strands. (Adopted from Jinek et al., 2012 and modified.)

4.2. Modification of the Cas9 nuclease domains

Sometimes, in order to edit the gene of interest with the help of the CRISPR/Cas9 technology, the creation of the DSB is not necessary. Thus, the Cas9 can be specifically modified not to cleave the target site at all or to nick only one of the DNA strands which enables the development of new efficient gene editing strategies.

By installing a specific H840A mutation in the HNH and a D10A mutation in the RuvC nuclease domains of Cas9, changing its properties from DNA-cleaving protein to a DNA-binding one can be achieved (Jinek *et al*., 2012). This new version of Cas9 is called deadCas9 (dCas9) and finds its purpose in multiple CRISPR/Cas9-mediated gene editing strategies. For instance, in the first and second generations of cytidine base editors (CBEs), dCas9 effectively leads the associated deaminase to the target site and, therefore, enables the creation of the desired edit (Komor *et al*., 2016).

By mutating only one of the nuclease domains, Cas9 nickase (Cas9n), able to cleave only one DNA strand at a time, is created. (Jinek *et al*., 2012). Using two Cas9 nickases to target the same genomic site but each nicking a different DNA strand immensely enhances the target specificity of the CRISPR/Cas9 system (Ran *et al*, 2013). Moreover, Cas9n mutants are routinely utilized in gene editing for their convenient attributes. Base editing, for example, takes advantage of the Cas9n D10A mutant cleaving only the complementary DNA strand (Komor *et al*., 2016). On the other hand, prime editing uses the Cas9n H840A mutant nicking only the non-complementary DNA strand to achieve efficient editing (Anzalone *et al*., 2019).

5. Genome editing methods and strategies

The choice of an appropriate gene editing approach to use in order to achieve the desired genome alteration is extremely important for the success of a given experiment. Numerous different methods employing the CRISPR/Cas9 technology exist, thus, enabling scientists to introduce various precise deletions, insertions, substitutions and other mutations into the genome.

5.1. Gene knock-out

Gene knock-out is one of the oldest and simplest editing strategies whose purpose is to inactivate the gene of interest connected to the subsequent depletion of its product (Duan *et al*., 2016). Many different approaches for inducing gene knock-out have been developed indicating its universal applicability in the scientific field.

5.1.1 Single gRNA strategy

The original strategy developed to inactivate genes using the CRISPR/Cas9 system takes advantage of the endogenous cellular NHEJ repair mechanism. After Cas9 is led to the target sequence

by associated sgRNA, it introduces a DSB into the genomic site which is predominantly repaired by this pathway (Jinek *et al*., 2012, Takata *et al*., 1998) However, NHEJ is error-prone and creates small insertions and deletions (indels) which usually cause a frameshift in the target protein-coding sequence and can result in the formation of premature STOP codons (Yen *et al*., 2014).

The choice of target site in the gene of interest is crucial for the efficiency of gene knock-out. The most successful approach seems to be targeting the critical exon of a given gene. Critical exon must be contained in all known splicing variants of a given gene in order for all of them to be affected. It also causes frameshift mutations when deleted, which potentially lead to the creation of premature termination codons and subsequent degradation of the mRNA (Bradley *et al*., 2012, Couttet and Grange, 2004). Finally, critical exon should be situated in the first 50 % of the coding area of the gene, as premature STOP codons in this region have a higher chance of being recognized, thus, marking the transcript as faulty (Hall and Thein, 1994).

The degradation of unusual mRNAs, for example, those with premature termination codons, is performed via the nonsense-mediated decay (NMD) pathway, one of the best characterized cellular quality control mechanisms. These flawed transcripts are recognized by the components of NMD, which then recruit other factors necessary for the pathway, and the damaged mRNA is degraded before it has a chance to be translated into protein. All these steps ensure the successful inactivation of the targeted genomic sequence (Couttet and Grange, 2004).

Duan *et al*. used the CRISPR/Cas9-mediated knockout strategy described above to generate a mouse model which could potentially help scientists develop an effective therapy to treat osteoporosis in humans. V-ATPase is an important and multifunctional enzyme and mutations in some of its subunits, for instance, ATP6V1H, are associated with various bone alterations. In this experiment, specifically designed sgRNA was utilized to lead Cas9 to the exon 2 of the *Atp6v1h* mouse gene where the nuclease created a DSB. The reparation of the cleavage by the imprecise NHEJ pathway resulted in the deletion of 5 nucleotides and a single-nucleotide substitution. These mutations then caused a frameshift and the formation of premature STOP codon leading to gene inactivation and depletion of the Atp6v1h subunit. Using this strategy, Duan *et al*. successfully created knock-out mouse models with decreased bone density and insufficient bone formation and calcification (Duan *et al*., 2016).

The CRISPR/Cas9 technology can also be employed to inactivate multiple genes simultaneously. Wang *et al*. transfected the Cas9 nuclease and a specific set of sgRNAs, each targeting a different gene, into host cells and observed knock-out of all targeted genes with high efficiency (Wang *et al*., 2013).

5.1.1.1. Possible challenges

Creating only one DSB is not always efficient enough to result in complete gene knock-out. The NHEJ pathway introduces indels into the genome in a rather random manner and the outcomes of this activity can differ tremendously (Yen *et al*., 2014). If the premature termination codon is situated in close proximity to the 5’ end of the mRNA, this transcript can escape the degradation mediated

by NMD by reinitiating the translation with the help of another START codon located downstream (Neu-Yilik *et al*., 2011, Zhang and Maquat, 1997). Random indels can also cause frameshift which affects the splice-regulatory sites of the targeted gene. Mutation of these sites might result in alternative splicing and, subsequently, in the production of aberrant proteins with unpredictable functions. They can display lower levels of their original activity or, sometimes, even negative dominant activity, which interferes with the usual function of the wild-type protein (Kapahnke *et al*., 2016).

5.1.2. Dual gRNA strategies

When targeting multiple genes, scientists discovered that the creation of two DSBs in adjacent genomic sites by sgRNA-associated Cas9 nuclease can cause the deletion of the intervening sequence (Cong *et al*., 2013, Zhou *et al*., 2014). This appeared to be an innovative approach for generating knock-out and research was conducted to further characterize this phenomenon, using numerous pairs of sgRNAs to target diverse genomic sites.

It was observed that the dual sgRNA strategy can delete sequences of various sizes from less than 1 kb to more than 1 Mb. Therefore, the excision of specific short sequences, exons as well as introns, or even whole genes is possible. Deletion of a critical exon appears to be a safer strategy for effective knock-out than deletion of the entire gene. The reason for this is that a gene can contain regulatory sequences responsible for modulating other non-targeted loci, which there is no desire to damage (Canver *et al*., 2014).

This CRISPR/Cas9-mediated deletion strategy was utilized by Syding *et al*. to create a mouse model of Angelman syndrome. This syndrome is a rare disorder characterized by intellectual and behavioural deficits and impaired motor skills. *UBE3A* gene, encoding the ubiquitin E3 ligase, is paternally imprinted in neurons leading to the predominant expression of the maternal copy. Angelman syndrome is usually caused by large deletions in this maternal copy resulting in depletion of the ligase which leads to the specific clinical manifestations. Syding and group employed dual gRNA to create a 76 kb deletion of the whole mouse *Ube3a* gene and generated models sharing many characteristics with patients suffering from this disease (Syding *et al*., 2022).

According to Canver *et al*., who studied the deletion properties of various genomic sequences, there is an inverse relationship between the size of the excised sequence and the frequency of such excision (Canver *et al*., 2014). Additionally, some studies show that the junction of DNA strands after the deletion of the intervening region was error-free and precise (Zheng *et al*., 2014) while others described the formation of various indels at the junction site as common (Canver *et al*., 2014).

This method can be utilized to delete coding as well as non-coding sequences, for example, long non-coding RNAs. Non-coding loci are not affected by frameshift mutations and, thus, cannot be inactivated by NHEJ-mediated incorporation of random indels. Therefore, deleting promoters of non-coding sequences or the excision of the entire sequence itself is an extremely convenient

approach for mediating their inactivation. This strategy can help researchers discover their real function and the exact way in which they affect phenotype (Han *et al*., 2014).

Additionally, it is rather difficult to predict the exact location of regulatory sequences needed for proper splicing. Unwanted damage in these sequences could result in alternative splicing and unpredictable products as described above. It is, therefore, advised to target the Cas9 cleavage to the sites at least 100 bp away from the boundaries between exons and introns (Miura *et al*., 2018).

Zuo *et al*. later observed that targeting a single critical exon of a gene by two or more closely spaced (from 10 to 200 bp apart) sgRNAs can achieve a rapid increase in knock-out efficiency. This enhancement seems to be caused by the simultaneous sequence deletion and indel formation leading to frameshift mutations. Their study also suggests that the more sgRNAs targeting the same exon were used, the larger deletions occurred (Zuo *et al*., 2017).

5.1.3. Double-nicking strategy

Using Cas9 nuclease to create DSBs can, sometimes, be connected to mutations in non-targeted genomic sites. These alterations of sequences with only partial homology to sgRNA spacer are possible due to the low number of base mismatches being tolerated (Fu *et al*., 2013). An increase in specificity is, thus, required for this method to yield satisfactory results. To overcome this challenge, a double-nicking strategy was developed. This approach uses Cas9n, with D10A mutation in its RuvC nuclease domain, altered to nick only one strand of the DNA. Furthermore, two sgRNAs are employed, each with a spacer complementary to one of the opposing DNA strands in the target site. These sgRNAs then lead the Cas9 nickases to create nicks in both strands at an appropriate distance from each other. This way a DSB is formed, which is then repaired via the NHEJ pathway in the same manner as those caused by the Cas9 cleavage. Double-nicking was reported to increase targeting specificity by up to 1,500-fold over wild-type Cas9 since two sgRNAs have to recognize the same target site for the DSB to be created. In addition, single-stranded nicks are usually repaired with high-fidelity, therefore, mutation rates are decreased in case of imprecise binding of only one of the sgRNAs into an off-target site (Ran *et al*., 2013).

Similarly as with the creation of a single DSB, the double-nicking strategy can also notably enhance the specificity of targeting and simultaneously decrease the occurrence of off-target effects while deleting the desired DNA sequence. Two pairs of sgRNAs are utilized to lead Cas9 nickases to two different target sites where they introduce nicks into both DNA strands as described above. This results in the formation of DSBs in both chosen sites and, subsequently, a possible deletion of the intervening sequence (Ran *et al*., 2013).

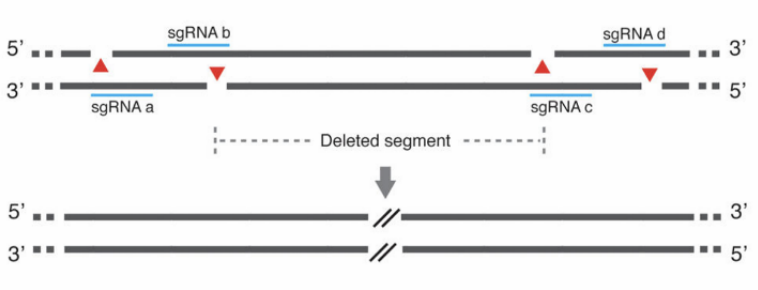

Figure 2: Use of double-nicking strategy to create genomic deletions
Two pairs of sgRNAs are delivered into the system, each targeting one side of the genomic site whose deletion is desired. The two sgRNAs in each pair bind to complementary DNA strands and lead Cas9 nickases to create single-strand nicks at an appropriate distance from one another, resulting in two DSBs. The region between those two DSBs is, subsequently, deleted. (Adopted from Ran et al., 2013 and modified.)

5.2. Gene knock-in

Gene knock-in is a precise method of gene editing which enables scientists to insert short as well as long exogenous sequences into the genome or replace existing endogenous sequences with new ones (Byrne *et al*., 2015, Raveux *et al*., 2017).

The original strategy for gene knock-in is based on the endogenous cellular HDR pathway (Raveux *et al*., 2017). After the DSB occurs in the genome, HDR can be activated to repair it through the process of homologous recombination. This mechanism requires a homologous sequence, often a sister chromatid, which it uses as a repair template to precisely fix the damaged DNA. HDR generally avoids the creation of additional mutations such as indels formed during the NHEJ reparation process (Wu *et al*., 2013).

To employ this pathway for the insertion of the desired exogenous sequences into the genome, scientists first used sgRNA-associated Cas9 nuclease to cleave the DNA and create a DSB. For initiation of HDR, the system must be supplied with a donor template which carries the desired insert flanked by arms homologous to the DNA sequence adjacent to the cleavage site. These arms then facilitate homologous recombination and incorporation of the insert into the genome (Raveux *et al*., 2017).

The template can be provided in the form of a double-stranded circular plasmid, double-stranded linear DNA sequence created by polymerase chain reaction (PCR) or single-stranded oligonucleotide (ssODN) (Raveux *et al*., 2017, Song and Stieger, 2017).

5.2.1. Homology-dependent gene knock-in

The classical HDR-dependent knock-in technique often utilizes the dsDNA plasmid as a donor and Raveux *et al*. successfully incorporated a gene fragment approximately 1 kb long with the help of homology arms the size of 500 bp using this template (Raveux *et al*., 2017).

Nutter *et al*. reported the creation of mouse models of myotonic dystrophy type 1 with this classical homology-dependent knock-in approach. This rare disease observed in humans is caused by CTG repeat expansion in the *DMPK* gene which results in progressive muscle weakness

and wasting (Brook *et al*., 1992). This group inserted a plasmid donor containing a large number of CTG repeats into the mouse *Dmpk* gene. Final models were confirmed to be valuable for testing drugs with the potential to treat muscular dystrophy type 1 (Nutter *et al*., 2019).

Usage of dsDNA plasmid is, however, connected to rather low editing efficiency. To address this issue, circular plasmids were linearized using restriction endonucleases or Cas9 cleavage. This alteration was reported to enhance the efficiency of HDR-mediated insertion in some studies (Song and Stieger, 2017) but decreased it in others (Beumer *et al*., 2008). Furthermore, the linearized plasmids were observed to increase the number of successfully modified target sites compared to linear dsDNA templates created via PCR. This is possibly due to the post-translational modifications or protection from degradation, both of which the PCR templates lack (Song and Stieger, 2017).

Moreover, multikilobase sequences can also be replaced via the HDR-based mechanism when the system is supplied with plasmid donors containing long homology arms. Interestingly, creating only one DSB in the target genomic location was observed to be more efficient compared to introducing two DSBs flanking the sequence to be replaced. The latter option often resulted in inversion or deletion of the intervening region, instead of its replacement (Byrne *et al*., 2015).

The desired insert might be also encoded by a synthetically generated ssODN donor. The length of commercially available ssODNs is, nevertheless, limited to about 200 bp. Thus, this method can be used to incorporate only short fragments of DNA using much shorter homology arms compared to the classical HDR-based strategy with plasmid templates. Raveux *et al*. and Miura *et al*. reported efficient incorporation of 30 bp and 40 bp long inserts with homology arms the size of 60 bp using this strategy. However, ssODN templates hold some advantages over classical dsDNA HDR-mediated insertion techniques. First, these templates are simple to design compared to plasmid vectors. Additionally, due to the smaller size of the vector, it is possible to transfer a higher number of donors into host cells which increases the editing efficiency. Finally, they are much less likely to be randomly incorporated into the genome compared to dsDNA vectors. Overall, the use of ssODN as a template was observed to be safer and more efficient than the use of double-stranded plasmids (Miura *et al*., 2018, Raveux *et al*., 2017).

To better understand the mechanisms accompanying osteogenesis imperfecta type V, Rauch *et al*. generated knock-in mouse models of this disorder using the ssODN donors. Osteogenesis imperfecta type V is a rare disease caused by a heterozygous single-nucleotide -14C<T mutation in the 5' untranslated region (UTR) of the *IFITM5* gene. This gene encodes the BRIL protein which is abundantly expressed in osteoblasts. This single-base substitution creates a novel translation initiation site, subsequently adding 5 amino acids to the N-terminus of the BRIL protein (Semler *et al*., 2012). Osteogenesis imperfecta type V is connected with impaired bone formation, low bone mass and other bone abnormalities. Rauch *et al*. designed sgRNA to target the 5' UTR of the mouse *Ifitm5* gene and inserted a 67-base-long ssODN sequence carrying the desired -14C>T mutation. The mosaic males obtained from this experiment were then mated with wild-type females resulting

in offspring which carried the desired heterogeneous mutation in the *Ifitm5* gene. These mice displayed a wide variety of bone abnormalities which unfortunately resulted in neonatal death (Rauch *et al*., 2018).

To bypass the size limitations of ssODN templates, the *in vitro* transcription and reverse transcription (ivTRT) technique was developed. This approach employs routine cellular processes to synthesize long ssODN donors. In the first step, a long dsDNA fragment containing the desired insert is transcribed *in vitro* to create RNA. Reverse transcriptase is then used to revert the RNA sequence to DNA resulting in final long ssODN templates (Miura *et al*., 2018).

The Efficient additions with ssDNA inserts-CRISPR (Easi-CRISPR) strategy utilizes these long ssODN donors to incorporate desired DNA fragments into the genome. The Cas9 protein along with separate crRNA and tracrRNA are delivered into host cells to create DSB at the target genomic site. This approach was observed to have enhanced efficiency compared to using Cas9 mRNA or Cas9 protein with sgRNA (Aida *et al*., 2015). Simultaneously, the cells are supplied with long ssODNs containing desired inserts flanked by homology arms. Successful incorporation of up to 1.4 kb long fragments with homology arms about 100 bp in size was recorded. This method is generally applicable as it was shown to successfully work for over a dozen different genome loci with rather high editing efficiencies (Miura *et al*., 2018, Quadros *et al*., 2017).

The CRISPR/Cas9-mediated Precise Integration into Target Chromosome (CRIS-PITCh) is a technique which employs the microhomology-mediated end joining (MMEJ) repair pathway for incorporation of desired edits into the genome. MMEJ is active during the G1 and early S phases of the cell cycle and, unlike HDR, requires only very short homology arms, ranging from about 5 to 40 bp, for efficient insertion of DNA fragments. This strategy, similarly to the Easi-CRISPR, shows the best results when crRNA and tracrRNA are not fused into a single sgRNA. The donor template is supplied in the form of a dsDNA circular plasmid containing the desired edit flanked by short homology arms. Additionally, two to three different crRNAs are used in this approach: one cleaving the target site in the genome and one or two more cleaving the target sites in the donor vector. When incorporation of the whole plasmid is desired, only one crRNA is used to cleave it, and the plasmid is linearized. Nevertheless, double cleavage of target sites adjacent to microhomology arms by two crRNAs releases the insertion cassette and rids it of unnecessary plasmid backbone. Successful incorporation of a large fragment the size of 5 kb was reported using homology arms of only 40 bp in length. Overall, the CRIS-PITCh is a flexible method of gene knock-in and shows higher efficiency than the classical HDR-mediated DNA fragment insertion (Aida *et al*., 2016, Nakade *et al*., 2014, Sakuma *et al*., 2016).

In another approach developed by Zhang and group, the desired insert is flanked by long homology arms and delivered into host cells in the form of a dsDNA circular plasmid. This donor also contains one additional sgRNA target site on each side of the insertion cassette, similarly as in the PITCh method. Along with the insert, the Cas9 mRNA and sgRNAs are co-transfected to cells, resulting in simultaneous cleavage of the genomic target site and both target sites in the donor vector.

The cassette released from the template is then incorporated into the genomic site. This strategy was successfully reported to insert a 2 kb long DNA fragment with the use of homology arms ranging from 600 to 1000 bp. It combines long homology arms from the classical HDR-based method and the design of the donor template from the MMEJ-based technique resulting in higher editing efficiency than both of these previous approaches (Zhang *et al*., 2017).

Targeted integration with linearized dsDNA-CRISPR (Tild-CRISPR) is a very similar strategy to the method developed by Zhang and group. The donor template is, as well, provided in the form of a dsDNA circular plasmid containing the desired insert flanked by long homology arms. However, the main difference is that in the Tild-CRISPR the donor vector is linearized *in vitro* using two restriction nucleases, not upon delivery by Cas9-mediated cleavage. The restriction enzymes recognize their target sites adjacent to each homology arm resulting in the release of the insertion cassette. This cassette is, subsequently, co-transfected into cells along with the CRISPR/Cas9 system components and incorporated into the cleaved genomic site. This technique can successfully insert fragments up to 6 kb long with homology arms the size of about 800 bp. The Tild-CRISPR was observed to create the desired edits with higher efficiency than the classical HDR-based method, the strategy developed by Zhang and even the Easi-CRISPR approach (Yao *et al*., 2018).

5.2.2. Homology-independent gene knock-in

Although the original approaches for DNA insertion are based on HDR or other homology-dependent pathways, the incorporation of gene fragments is also possible via NHEJ. Furthermore, as the NHEJ pathway is active during the whole cell cycle, unlike HDR (Takata *et al.*, 1998), it has been reported to insert desired edits with higher frequency than the classical homology-based strategy. This approach was observed to incorporate extremely long fragments of up to 34 kb into the genome without the need for homology arms. It is, nevertheless, important to note that using larger donors immensely decreases the knock-in efficiency. NHEJ-based incorporation employs donor templates in the form of dsDNA circular plasmids linearized by the Cas9 cleavage upon co-transfection inside cells. The target gene is cleaved by Cas9, as well, and the insert is incorporated into the genome by the direct joining of blunt DNA ends. The donor template can carry either one sgRNA target site on one side of the insert (single-cut donor) or two sgRNA target sites flanking the insert on both sides (double-cut donor). The double-cut donor strategy creates two linear fragments, one with the desired edit and one without it. These two fragments then compete for the integration into the genome and decrease the efficiency of insertion. Using a single-cut donor approach is, thus, a more convenient option. The disadvantage of this technique is that due to the direct joining of blunt ends, the insert can be incorporated in both directions. In addition, there is a higher chance of incorporation of the DNA fragments into off-target sites compared to HDR-based methods

as this strategy does not require any homology in the donor template. The formation of various indels is also much more frequent compared to homology-dependent knock-in approaches due to the nature of the NHEJ repair pathway (He *et al*., 2016).

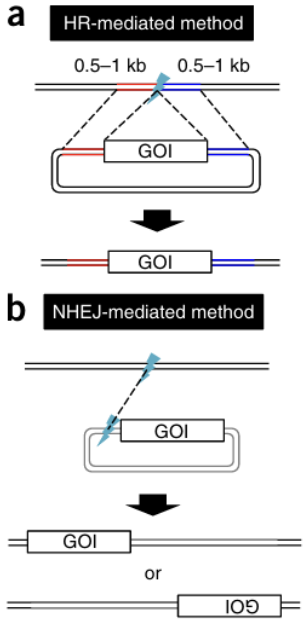

Figure 3: Schematic of homology-dependent and homology-independent knock-in methods
A) The homology-dependent knock-in utilizes homologous arms of specific length flanking the desired insert in the donor vector. This insert is integrated into the target genomic site previously cleaved by Cas9 nuclease through the process of homologous recombination. B) In the homology-independent knock-in, the donor vector carrying the desired insert is cleaved by Cas9 and integrated into the target genomic site also previously cleaved by Cas9. This method does not require any homologous arms, however, the insert can be integrated in random orientation. The blue lightnings symbolize the Cas9 cleavage sites, the blue and red sequences symbolize the homology arms, GOI: gene of interest. (Adopted from Sakuma et al., 2016 and modified).

An alternative technique of NHEJ-mediated insertion employs Cas9 nickases which are used to nick both strands of DNA with a precise offset and, therefore, create a DSB with specific 3' or 5' overhangs. If the system is simultaneously supplied with double-stranded oligonucleotides with matching overhangs, these templates can be incorporated into the genome via non-homologous ligation (Ran *et al*., 2013).

The Knock-in blunt ligation (KiBL) strategy uses a linear dsDNA fragment created by PCR as a donor template. This fragment is co-transfected into cells along with the CRISPR/Cas9 technology components and is incorporated into the target genomic site cleaved by Cas9. This method has been observed to efficiently knock-in gene fragments of up to 2 kb in size. Nevertheless, the length limitations of linear dsDNA sequences that can be generated by PCR, the low efficiency of electroporation of these donors into cells and, finally, the possibility of the insert being incorporated in the wrong direction are all drawbacks scientists have to face when choosing this method for their experiments (Geisinger *et al*., 2016).

The CRISPR/Cas9 homology-independent targeted insertion (CRISPR-HITI) is another knock-in technique completely independent of homology arms. Its main advantage is the ability to bypass the incorporation of gene fragments in random directions. The desired insert is encoded in dsDNA circular plasmid which is, upon transfection into cells, cleaved in one or two target sites and linearized. The same sgRNA used for the Cas9-mediated cleavage of the vector also targets the chosen site in the genome. However, the orientation of the target site in the donor template is flipped compared to the one in the genome. This modification ensures that the DNA fragment is inserted in the desired orientation as the reverse orientation results in recreating the intact sgRNA target site. This intact target site is then repeatedly cleaved until the right orientation of the insert is achieved or the target site is damaged by the formation of random indels. The knock-in efficiency of the CRISPR-HITI strategy was observed to be higher than that of the classical HDR approach or the PITCh method (Suzuki *et al*., 2016).

The Targeted Knock-In with Two guides (TKIT) is a specific technique which mediates the incorporation of the exogenous DNA sequence into the non-coding areas of the targeted gene. It employs a pair of sgRNAs to guide Cas9 to cleave two sites inside the targeted intron. The same pair is then used for the Cas9-mediated cleavage of two sites flanking the desired insert in the dsDNA circular donor plasmid. The target sites in the vector are, nevertheless, flipped and switched compared to the ones in the genome to ensure the right orientation of fragment incorporation. Similarly as in the CRISPR-HITI approach, the sequence inserted in the reverse direction recreates the intact sgRNA target site. This site is then cleaved by Cas9 until the fragment incorporates in the forward orientation or until the site is destroyed by indels. Due to the targeting of non-coding sequences, no exons in the gene of interest are cut, which makes this method less sensitive to indel mutations. TKIT has been reported to have higher knock-in efficiencies compared to the CRISPR-HITI approach (Fang *et al*., 2021).

5.2.3. Transgenic mice

As previously described, the CRISPR/Cas9 system can be utilized to insert rather long DNA fragments, even whole genes, into the mouse genome. By incorporating exogenous genes originating from different organisms, transgenic mice are created. These transgenes then have the potential to be expressed and produce RNA or even be translated into proteins (Chu *et al*., 2016).

The insertion of transgenes has been frequently targeted into a specific locus in the mouse genome. This locus located on chromosome 6 is termed ROSA26 and is expressed in all cell types as well as all stages of development making it a convenient candidate for transgene integration. This locus consists of three exons and produces three different transcripts. Nevertheless, these transcripts are not translated into any working proteins and their function is not yet clarified.

The transgene insertion into ROSA26 is, thus, relatively safe as the possible damage caused by imprecise knock-in should not lead to any changes in the mouse phenotype. The insert is usually targeted into the first intron of the ROSA26 locus. It contains either a splice acceptor which ensures its fusion with the first exon mediated by cellular splicing machinery, or a CAG promoter allowing it to be expressed independently of the locus (Soriano, 1999, Zambrowicz *et al*., 1997).

Humanized mice, with the inserted exogenous DNA sequences originating from humans, form a specific subtype of transgenic mice. Such models are priceless for medical research as they bring the experimental system closer to real conditions inside the human body. Even though mice were selected as the fittest organism for modelling human diseases, their anatomy, physiology and genetic information are still very different from humans. Humanizing mice models by insertion of human genomic sequences or engraftment of human cells makes the experimental results more translatable. This approach is extremely useful for the development of therapeutical approaches to treat various human diseases (Zhang Y. *et al*., 2022).

Humanization of even very long genomic sequences is possible as shown by Leidy-Davis *et al*. who reported successful replacement of the 17 kb *Bcl2l11* mouse gene with the 25 kb long homologous human *BCL2L11* gene. In this experiment, the desired insert was encoded by a circular donor vector which was integrated into the target genomic site via homologous recombination. Additionally, two pairs of sgRNAs were used, each of them targeting one side of the mouse *Bcl2l11* gene, to ensure the efficiency of creating a DSB and its longer persistence (Leidy-Davis *et al*., 2018).

To enhance the development of appropriate therapy for rare diseases, Zhang Y. *et al*. successfully employed the CRISPR/Cas9 technology to create a humanized mouse model of a lethal neuromuscular disorder called Duchene muscular dystrophy. This rare disease is caused by various mutations in the *DMD* gene encoding protein dystrophin which is essential for effective muscle functioning (Koenig *et al*., 1987). The paper describes the replacement of mouse exon 51 of the *Dmd* gene with the 233 bp long exon 51 of the homologous human *DMD* gene via CRISPR/Cas9-mediated homology-dependent knock-in. The CRISPR/Cas9 system was then used to delete the mouse endogenous exon 50 leading to frameshift and creation of premature STOP codon in human exon 51. This resulted in inactivation of the *Dmd* gene and prevented the expression of dystrophin protein causing Duchenne muscular dystrophy in the humanized mice model. Furthermore, they managed to correct this mutation by creating a single DSB with sgRNA-associated Cas9 between the splice acceptor and the premature termination codon in the exon 51. The NHEJ-mediated repair and formation of specific indels resulted in the correction of the open reading frame, which ultimately led to the restoration of dystrophin production. It was also observed that large indels disrupted the splice acceptor and caused the skipping of the damaged exon 51 (Zhang Y. *et al*., 2022).

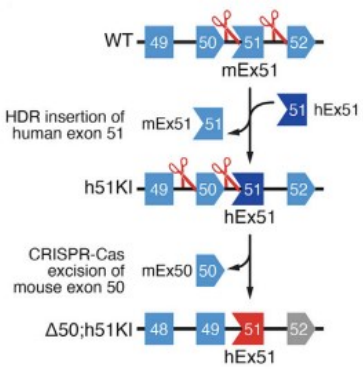

Figure 6: Modeling of the Duchenne muscular dystrophy humanized mouse model
The CRISPR/Cas9 system was utilized to exchange the exon 51 of the mouse *Dmd* gene for the exon 51 of the homologous human *DMD* gene using a homology-dependent knock-in strategy. Subsequently, CRISPR/Cas9-mediated deletion of the endogenous mouse exon 50 caused a shift of reading frame and formation of the premature STOP codon in the human exon 51 resulting in *Dmd* inactivation. The red scissors symbolize the Cas9 cleavage sites. (Adopted from Zhang Y. et al., 2022.)

5.2.4. Gene knock-out mediated by homology-directed repair

Although NHEJ is the original and most often used pathway for the production of knock-outs, HDR can be utilized to inactivate genes as well by inserting specific sequences into their critical exon. The first option is the incorporation of small sequences, such as specific restriction sites, which then cause frameshift mutations and the creation of premature STOP codon (Heeb *et al*., 2023). Another strategy is the insertion of a larger sequence, for example, a whole reporter gene along with its polyadenylic acid (poly(A)) tail. This tail ensures the premature termination of the targeted gene transcription, resulting in its inactivation. Additionally, the reporter gene facilitates the selection of successfully mutated cells (Zare *et al*., 2018). Finally, gene trapping is a method in which a whole trapping cassette is integrated into the first intron of the gene of interest. This cassette usually consists of a 3' splice site, selection marker gene (for example lacZ) and poly(A) tail. As in the previous strategy, transcription of the targeted gene is terminated prematurely due to the poly(A) sequence and the targeted gene is, therefore, knocked out. Moreover, the 3' splice site ensures the fusion of the first exon and selection marker gene which is then expressed and once again simplifies the process of identifying correctly modified genes (Reber *et al*., 2018).

5.2.5. Conditional knock-out

Conventional knock-out of essential genes, such as housekeeping genes, is connected to prenatal lethality or premature postnatal death, which poses a challenge for scientists attempting to study their properties (Lindhurst *et al*., 2006, Luo *et al.,* 1997). In addition, it is rather convenient to be able to study the effects of gene inactivation only in specific cell types or tissues as the outcomes of such site-specific disruptions can differ tremendously (Nakagawa *et al*., 2016). A simple knock-out strategy, however, cannot direct the gene inactivation only in specific areas of the mouse body, making researching this phenomenon further rather problematic. A conditional knock-out is an approach which solves

these difficulties by allowing scientists to spatiotemporally control gene inactivation with the help of inducible or site-specific recombinases (Akagi, *et al*., 1997, Feil *et al*., 1996, Gossen *et al*., 1995).

The cyclization recombinase (Cre) which recognizes specific sites termed locus of crossing over, P1 (loxP) is the most often used recombinase in conditional knock-out experiments. It can mediate recombination between two adjacent loxP sites and effectively delete the intervening sequence in the process. Thus, by precise positioning of loxP sites inside the targeted gene, the creation of a so-called conditional or floxed allele is possible. Cre then initiates the recombination and subsequent excision of the critical sequence resulting in gene inactivation and leaving only one of the loxP sites (Nakagawa *et al*., 2016, Yang *et al*., 2013). The same principle applies to flippase (FLP) recombinase which recognizes the FLP recognition target (FRT) and is also frequently used in conditional knock-out approach (Chen *et al*., 2018).

The inducible recombinases can be activated in any chosen stage of life by adding an exogenous chemical to the conditional knock-out system. The most frequently used chemicals are tetracycline or its derivatives, such as doxycycline, and tamoxifen. Tetracycline is able to induce Cre or FLP expression on the transcriptional level while tamoxifen directs the activity of Cre into the nucleus on the protein level. (Feil *et al*., 1996, Gossen *et al*., 1995).

To promote a tissue-specific knock-out, the recombinase gene must be positioned under the control of the tissue-specific promoter. Each of these promoters displays its activity only in a particular cell type inside the model organism enabling scientists to direct the knock-out to desired areas, for example, liver, retina or Schwann cells (Akagi *et al*., 1997).

The main advantage of this strategy is that the genes are not inactivated from the moment the changes in the genome are executed. This is possible because conditional alleles are specifically designed not to interfere with normal gene expression. Recombinase recognition sites are situated inside introns or even the coding exons, but in both situations spliced out of the final transcript (Wu *et al*., 2022, Yang *et al*., 2013). Therefore, the targeted gene is fully functional until the recombinase is expressed leading to gene inactivation.

By incorporating the recombinase recognition sites into their genome, a line of mice containing conditional alleles can be obtained. As they do not naturally express Cre, another line of transgenic mice with inducible or site-specific recombinase must be created. These two lines then need to be bred, eventually creating offspring carrying a conditional allele as well as a gene for Cre. Only then are the mice ready for conditional knock-out experiments (Wu *et al*., 2022).

5.2.5.1. Creation of conditional allele

There are numerous strategies for incorporating the desired sequence into the genome to make the subsequent conditional knock-out possible.

HDR-based insertion of recombinase recognition sites on each side of a critical exon is one way to create a conditional allele. The insert can be encoded by a classical circular dsDNA vector

or by ssODNs with shorter homology arms (Chen *et al.,* 2018, Yang *et al*, 2013). Another knock-in strategy, often used to create floxed alleles, is the Easi-CRISPR which employs long single-stranded DNA donors carrying the whole critical exon flanked with recombinase recognition sites (Quadros *et al*., 2017). Finally, the PITCh method encodes the desired insert in the form of circular dsDNA. This approach, however, takes advantage of MMEJ instead of HDR (Aida *et al*., 2016).

Because the conventional knock-out of the mouse *Mgp* gene encoding Matrix Gla protein (MGP) is lethal, Borrás *et al.* created the *Mgp* floxed mouse model using the CRISPR/Cas9 technology. Mutation in the human *MGP* gene is associated with Keutel syndrome which is a rare disease characterized by abnormal calcification in the lung, cartilage and vascular system (Munroe *et al*., 1999). This group employed CRISPR/Cas9 to insert a single-stranded circular donor vector into the mouse *Mgp* gene. This donor contained a specific part of the gene flanked by precisely located loxP sites and was used to replace the corresponding part of the endogenous DNA sequence to enable conditional knock-out. As high expression of *Mgp* was observed in the trabecular meshwork of the eye, Cre was specifically transferred into this tissue through adenoviral-mediated delivery. Cre then performed successful recombination between the two loxP sites leading to *Mgp* knock-out, which caused elevation of intraocular pressure, one of the main risk factors for the development of glaucoma (Borrás *et al*., 2020).

The recombinase recognition sites can be inserted sequentially on one side of the exon and then the other or simultaneously on both sides. Nevertheless, a high frequency of chromosomal deletions was reported during concurrent insertion as it requires the formation of two DSBs at the same time. Sequential insertion, where the first recombinase recognition site is inserted into a 1-cell stage zygote and the second one into a 2-cell stage zygote, is, thus, a much safer option (Horii *et al*., 2017).

To study the effects of the *Creb3l3* gene knock-out in specific tissues, Nakagawa *et al*. generated floxed mouse models by simultaneously inserting recombinase recognition sites into the targeted gene using the CRISPR/Cas9 system. Subsequently, gene inactivation was induced either in the small intestine or in the liver. These two sites were specifically chosen since *Creb3l3*, which encodes an important regulator of lipid and glucose metabolism, is usually expressed in them. The floxed mice were created by concurrent cleavage in introns 3 and 11 of the *Creb3l3* mouse gene by two sgRNA-associated Cas9 nucleases and the insertion of one loxP site into each of the disrupted sites. The inserts were encoded by ssODNs and flanked by 250 bp long homology arms. As mentioned above, the simultaneous cleavage unfortunately resulted in a high frequency of unwanted deletions. However, the desired floxed mice were obtained as well enabling the group to continue with the experiment. These mice were then mated with transgenic mice carrying the Cre gene, whose expression was under the control of either liver- or small intestine-specific promoters. Mice with small intestine-specific knock-out displayed no obvious change in phenotype compared to floxed mice, nevertheless, mice with the liver-specific gene inactivation showed quite a few. They demonstrated hypertriglyceridemia and hypercholesterolemia as well as acceleration in the development

of non-alcoholic steatohepatitis, when fed the methionine-choline deficient diet. These conditions are all rare medical disorders observed in humans, which makes these mouse models quite useful in medical research (Nakagawa *et al*., 2016).

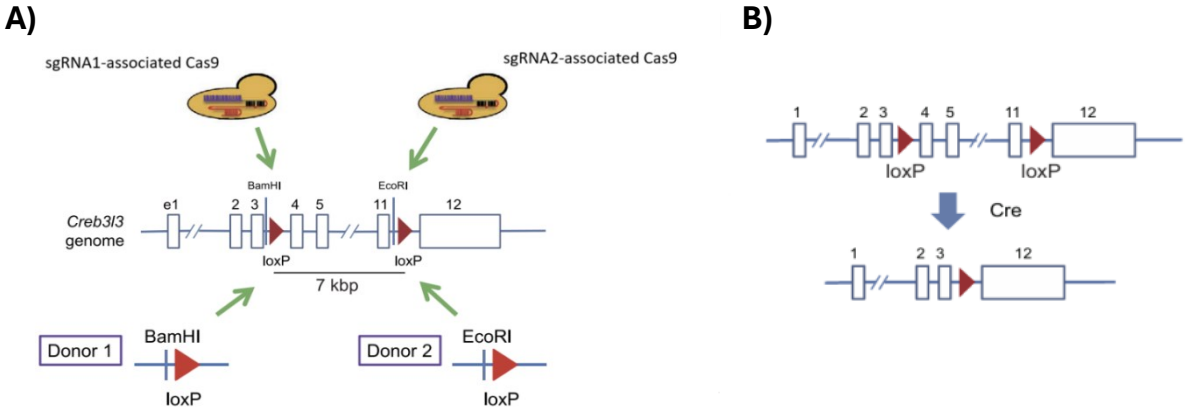

Figure 4: Creation of floxed allele and conditional knock-out by activation of Cre expression
A) Two sgRNAs target specific DNA sites inside introns of the *Creb3l3* mouse gene and guide Cas9 proteins to create DSBs. Oligonucleotides carrying loxP sites are inserted into the two cleaved sequences resulting in the creation of a floxed allele. B) The sequence between two loxP sites is deleted, after the tissue-specific expression of Cre, and the gene is inactivated. (Adopted from Nakagawa et al., 2016 and modified.)

Another intricate strategy used for the creation of conditional alleles is the HDR-mediated insertion of Short Conditional intrON (SCON) into the critical exon sequence. SCON is exactly 189 bp long and consists of a 5' splice donor site, followed by loxP, a branch point, another loxP, a polypyrimidine tract and, finally, a 3' splice acceptor site. Upon insertion, expression of the target gene is not affected by incorporated SCON, because the 3' and 5' splice sites ensure its excision from the transcript before translation takes place. However, when Cre is expressed, it mediates recombination between the two loxP sites, effectively deleting the branch point and shortening the SCON to 55 bp. Without the branch point, the SCON cannot be spliced out of the transcript and, moreover, the shortened version contains a termination codon in each of the three possible reading frames. This leads to premature termination of translation, NMD-mediated degradation of mRNA and, eventually, gene knock-out (Wu *et al*., 2022).

Finally, the knock-out first allele technique combines the possibilities of creating a simple conventional knock-out as well as a conditional knock-out. In this approach, a specifically constructed cassette is inserted into the intron adjacent to the critical exon. The cassette is referred to as tm1a and consists of a 5' FRT followed by a lacZ gene and a loxP site, the neomycin phosphotransferase gene and, at last, a second FRT followed by a second loxP site. The lacZ gene along with the neomycin phosphotransferase gene make the selection of clones with correctly installed incorporation easier. The cassette also contains a splice acceptor on its 5' end and a poly(A)sequence on its 3' end. After the insertion of a simple loxP recognition site into the intron on the other side of the critical exon, the knock-out first allele is ready (Testa *et al*., 2004).

The incorporation of the tm1a alone induces gene knock-out as it traps the gene by premature termination of the transcription due to the poly(A) tail. This inactivated gene can be then modified in two different ways. By inducing the expression of Cre, the neomycin phosphotransferase gene as well as the critical exon are excised, leaving only the lacZ gene flanked by one FRT and one loxP site. This way, the tm1b allele is created and the targeted gene remains knocked out. Another option is promoting the expression of FLP recombinase first. FLP recombinase mediates the recombination between FRP recognition sites and the deletion of both the lacZ and the neomycin phosphotransferase genes. The newly formed allele is termed tm1c and consists of two loxP sites flanking the critical exon and one FRT left after the FLP-mediated recombination. The activity of the previously knocked-out gene is restored, due to the deletion of poly(A) tail, and a floxed allele is created. After the subsequent activation of Cre, the critical exon is excised. This results in the final gene knock-out and creation of the tm1d allele with only one residual loxP and one FRT site remaining (Skarnes *et al.*, 2011, Testa *et al.*, 2004).

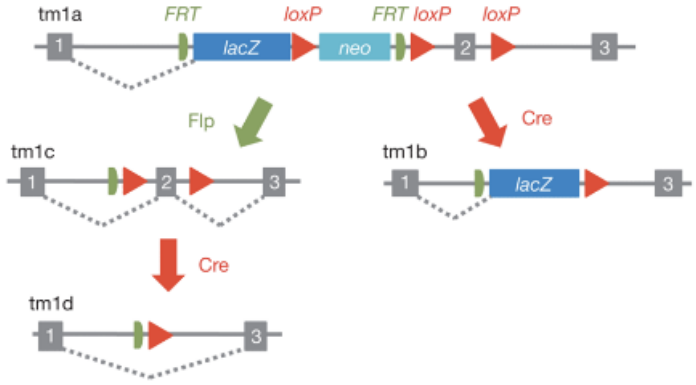

Figure 5: Schematic of the knock-out first allele
The insertion of the tm1a cassette traps the targeted gene and causes its inactivation. The expression of the FLP recombinase leads to the tm1a allele transformation into a conditional tm1c allele while the gene regains its function. However, the subsequent activation of Cre results in final gene knock-out due to the critical exon deletion. The tm1a allele can also be converted into the lacZ-tagged tm1b allele by expressing Cre first. In this allele, the critical exon is deleted as well. (Adopted from Skarnes et al., 2011.)

5.3. Base editing

Base editing is a novel method of utilizing the CRISPR/Cas9 system to generate single-nucleotide edits in a precise manner. Since the formation of DSB is not required in this approach, the Cas9 nuclease is mutated in either one or both of its nuclease domains, which results in creating Cas9n or dCas9, respectively. Modified Cas9 is fused with a specific base editor and led to the target sequence by sgRNA, where it can install the desired alteration into the genome (Komor *et al.*, 2016). Base editing works effectively in both dividing and non-dividing cells, unlike the HDR pathway (Yeh *et al.*, 2018).

Additionally, as no double-strand breaks are created, this strategy generates much cleaner products than the NHEJ pathway does (Komor *et al*., 2016). This strategy has remarkable potential in creating animal models of various rare diseases (Liang *et al*., 2017, Liang *et al*., 2018).

5.3.1. Base editors

Each base editing experiment is unique and, as such, each may require a different kind of substitution. Therefore, scientists developed multiple classes of base editors using naturally occurring or synthetic deaminases (Komor *et al*., 2016, Gaudelli *et al*., 2017, Xie *et al*., 2020, Kurt *et al*., 2021). For example, cytidine and adenine base editors quickly became popular among researchers for their convenient attributes and belong to the most frequently used editor systems today (Komor *et al*., 2016, Gaudelli *et al*., 2017).

5.3.1.1. Cytidine base editors

Cytidine deaminases catalyse the conversion of C into U, which is later substituted by T by the DNA replication process or cellular repair mechanisms. This eventually results in the change from the original GC pair to the novel AT pair (Komor *et al*., 2016).

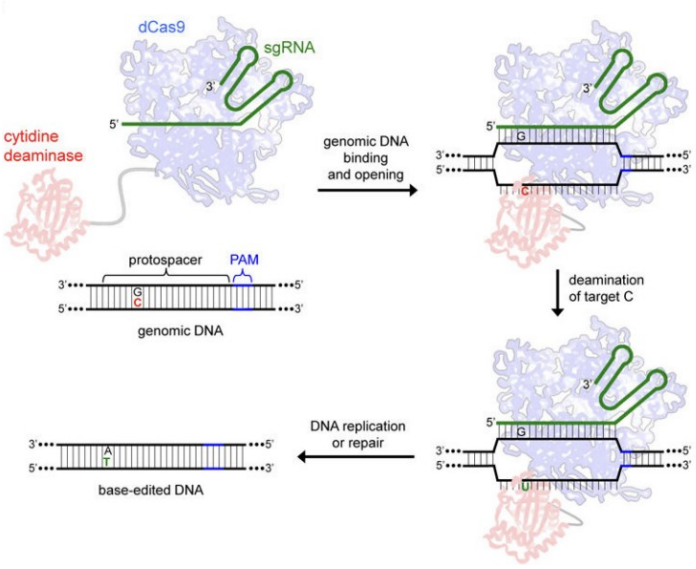

Figure 7: The C to T conversion mediated by cytidine base editors
The cytidine deaminase connected to catalytically inactive dCas9 is guided to the target genomic site by associated sgRNA and converts C into U. Subsequently, U is substituted with T in a process of DNA replication or repair and the former GC pair is turned into an AT pair. (Adopted from Komor et al., 2016.)

The apolipoprotein B mRNA editing enzyme, catalytic polypeptide (APOBEC) family of cytidine deaminases and activation-induced cytidine deaminase (AID) are naturally occurring proteins, able to edit ssDNA, employed to generate cytidine base editors (CBEs) (Komor *et al*., 2016, Nishida *et al*., 2016).

The first generation of cytidine base editors (CBE1) was created by Komor *et al.* It consisted of rat-derived cytidine deaminase (rAPOBEC1) connected to the N-terminus of dCas9 by 16 nt long XTEN linker and a nuclear localization signal (NLS) fused with its C-terminus. Although this system turned out to be rather successful in converting C to U during *in vitro* experiments, it performed poorly when introduced to human cells (Komor *et al*., 2016). The reason for this decrease in efficiency seemed to be the recognition of U by uracil N-glycosylase (UNG). This enzyme cleaves the glycosidic bond binding uracil to the DNA backbone, which stimulates the base excision repair (BER) pathway. The abasic site is, thus, repaired according to the DNA strand complementary to the sgRNA resulting in the original CG pair (Dianov *et al*., 1992, Lindahl, 1974).

Requiring a solution for this difficulty, scientists developed the second generation of cytidine base editors (CBE2) by adding a uracil DNA glycosylase inhibitor (UGI) to the C-terminus of CBE1. This adaptation inhibits UNG and, consequently, the BER pathway, resulting in an enhanced editing efficiency in human cells (Komor *et al*., 2016).

To create a mouse model of albinism, Liang *et al*. generated high-fidelity CBE2 by introducing five specific point mutations into the dCas9. Albinism is a rare disease caused by mutations in the *TYR* gene encoding an enzyme called tyrosinase which is necessary for melanin production (Körner and Pawelek, 1982). These mutations lead to the depletion of melanin and subsequent changes in phenotype, mainly loss of pigmentation. The C to T base conversion was targeted into the mouse *Tyr* and resulted in the creation of a premature termination codon which ultimately inactivated the gene. Therefore, tyrosinase could not be properly expressed and the desired mouse model was successfully generated (Liang *et al*., 2017).

Next, the catalytic activity of the HNH nuclease domain of dCas9 was restored. The third generation of cytidine base editors (CBE3), thus, contains a Cas9n with only the D10A mutation in its RuvC domain. The nick in the non-complementary DNA strand induced by Cas9n shifts the repair pathway to preferentially use the edited complementary strand as a template. As a result, an AU pair is created, which later undergoes the conversion into an AT pair by DNA replication machinery. Developing CBE3 enhanced the efficiency of editing two- to sixfold over CBE2, although a slight increase in indels was recorded as well (Komor *et al*., 2016).

The fourth generation of cytidine base editors (CBE4) was created by adding one more UGI to the N-terminus, to further prevent the BER pathway, and by extending the linker connecting rAPOBEC1 to Cas9n (Komor *et al*., 2017). Moreover, this system can be enhanced by replacing the original NLS with a bipartite NLS and attaching it to both the C- and N- terminus. This bis-bipartite NLS system then undergoes modification of codon usage, which ultimately results in an extremely efficient BE4max (Koblan *et al*., 2018).

The third and fourth generations of base editors were both additionally modified by fusion with Gam protein. This protein, originating from MU bacteriophage, reduced the number of indels in the final products. It has been proposed that mutations of this kind often form as a result of DSBs

created by simultaneous nicking of the non-complementary strand by Cas9n and the complementary strand by apurinic or apyrimidinic site (AP) lyase. After UNG cleaves the glycosidic bond of uracil, AP lyase is stimulated to nick the DNA strand adjacent to this novel abasic site (Komor *et al*., 2017). Gam protein binds to the DSB and prevents the NHEJ repair pathway, which could form undesired indels in the target DNA sequence (d'Adda di Fagagna *et al*., 2003).

Another cytidine base editor system utilizing AID, instead of APOBEC, was termed Target-AID. It consists of an AID ortholog called *Petromyzon marinus* cytidine deaminase 1 (pmCDA1) fused with the C-terminus of nCas9 along with additional UGI to prevent the BER pathway and increase C to T editing efficiency (Nishida *et al*., 2016).

5.3.1.2. Adenine base editors

There are no naturally occurring adenine deaminases converting bases in DNA templates. Therefore, the optimized TadA* deaminase had to be created in a laboratory using extensive engineering methods. It originated from tRNA adenosine deaminase A (TadA) which initiates adenine deamination in tRNA (Kim *et al*., 2006). The enhanced TadA* is able to deaminate A in the DNA and change it into I, which is later substituted by G with the help of the DNA replication process and cellular repair mechanisms. This leads to the ultimate replacement of the previous AT pair with the new GC pair (Gaudelli *et al*., 2017).

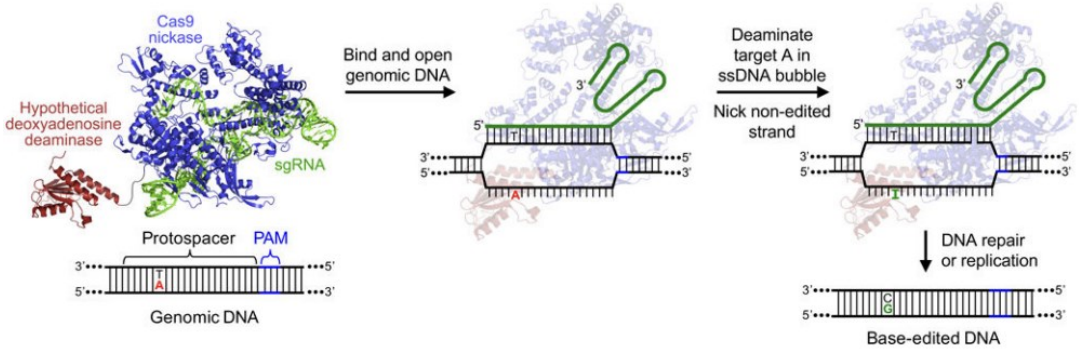

Figure 8: The A to G conversion mediated by adenine base editors
The sgRNA leads the Cas9 nickase and associated adenine deaminase to the genomic target site. There, the deaminase converts A into T and the Cas9 nicks the non-edited strand enhancing the editing efficiency. The edited T is then changed into G by DNA repair mechanisms or replication machinery. The former AT pair is substituted with a CG pair. (Adopted from Gaudelli et al., 2017.)

Adenine base editors (ABEs) yield final products that are significantly cleaner compared to CBEs. The reason for this is that the targeting and excision of converted inosine by alkyl adenine DNA glycosylase (AAG) (Lau *et al*., 2000) does not seem to decrease the editing efficiency, unlike the activity of UNG in CBEs (Gaudelli *et al*., 2017).

The first generation of adenine base editors, particularly ABE 1.2, has a similar constitution as the corresponding generation of CBEs. TadA* is connected to the N-terminus of nCas9 via XTEN linker while the C-terminus is fused with NLS. The observed editing efficiency of this generation was, however, very low (Gaudelli *et al*., 2017).

To overcome this issue, Gaudelli *et al.* introduced various novel modifications to the first generation of ABEs. One of them was the creation of a heterodimer consisting of wild-type TadA monomer and evolved TadA* monomer, which was then fused with nCas9. This way, new generations of ABE, including ABE7.10., were developed (Gaudelli *et al*., 2017). Furthermore, the replacement of the original NLS of the ABE7.10 system with bipartite NLS and attaching it to both termini, similarly as in CBE4, resulted in the creation of enhanced ABEmax (Koblan *et al*., 2018). Finally, ABE8e was established, which contained only the evolved TadA* monomer but with eight additional mutations, compared to ABE7.10, and displayed maximized editing efficiency (Richter *et al*., 2020).

ABE7.10 was successfully utilized by Liang *et al*. to target mRNA splice sites in the genes of interest and install A to G conversions. This effectively disrupted the process of correct gene splicing and resulted in the depletion of corresponding protein products. Using this strategy, they managed to create mouse models of two different rare diseases: albinism and Duchenne muscular dystrophy. In the first case, a single sgRNA was designed to target the splice donor at exon 3 of the *Tyr* gene where the ABE7.10 installed the desired edit. Unfortunately, no white-coated F0 mice were created. Nevertheless, when they were mated with homozygous *Tyr* mutant mice, albino pups were successfully obtained. When creating a Duchenne muscular dystrophy mouse model, two sgRNAs were designed to target the splice donor of exon 61 and splice acceptor of exon 66 in the mouse *Dmd* gene. These exons were chosen primarily because both of them are highly conserved between mice and humans. ABE7.10 system managed to convert A to G at the chosen splice sites with high efficiency and edited mice displayed phenotype typical for Duchenne muscular dystrophy in humans (Liang *et al*., 2018).

5.3.1.3. C to G base editor 1

Although cytidine deaminases used in CBEs typically initiate the gradual change of C into T, scientists discovered that they can be used to mediate the substitution of C by G as well. To establish a system facilitating this substitution, Kurt *et al*. decided to modify BE4max by removing the two tethered UGIs. As a result, the efficiency of C into G conversion was increased over the original BE4max, although a higher frequency of indels was also reported. Next, they decided to add human UNG protein to generate more abasic sites and activate the BER pathway. They hypothesized that this modification would enhance the editing efficiency, however, the frequency of base conversion was unexpectedly decreased. In order to further improve the properties of this complex, the R33A mutation was introduced to rAPOBEC1, which proved to be a successful approach as it considerably amplified the editing activity. The last step was the replacement of the human UNG for UNG extracted

from *E. coli* and fusing it with the N-terminus of the BE4max complex, which finally resulted in the creation of the C to G base editor 1 (CGBE1). The majority of edits installed by CGBE1 was observed in the cytidines situated in AT-rich sequence and those located at position 6 of the protospacer, when the end of the protospacer distal to PAM is counted as position 1 (Kurt *et al*., 2021).

During the process of CGBE1-mediated base substitution, a higher frequency of indel formation was reported compared to editing with BE4max. To solve this difficulty, the miniCGBE1 was created by removing the *E. coli* UNG protein from the CGBE1 complex. This system displayed a slight decrease in editing activity, nevertheless, fewer indels were observed in its final products (Kurt *et al*., 2021).

5.3.1.4. Adenine and cytosine base editor

By fusing both cytidine and adenine deaminases to Cas9n, adenine and cytosine base editor (ACBE) was developed, which are able to simultaneously convert C to T and A to G in the target region. Xie *et al*. chose a cytidine base editor Target-AID and an adenine base editor ABE7.10 for their unique complementary properties to create this new base-editing system. The PmCDA1 along with the UGI protein of Target-AID were fused to the C-terminus of ABE7.10 resulting in a powerful C to T and A to G converter. Moreover, it was observed that the length of the sgRNA spacer as well as the length of the linkers in the ACBE system have a strong impact on the editing efficiency. According to the study, the sgRNA spacer should be ideally 20 nt long and the linker connecting nCas9 and TadA heterodimer of ABE7.10 should contain 16 amino acids to achieve optimal dual base editing (Xie *et al*., 2020).

5.3.2. Editing window

After the base editor system binds to its target sequence, an R-loop is formed displacing the non-targeted strand of the DNA. This temporarily creates a single-stranded area in the complementary strand, part of which is available for base conversion and is called an editing window. The editing window of CBE3 and ABE7.10 ranges from approximately 4 to 8 bases and from 4 to 7 bases of protospacer, respectively. As the deaminases have access to all of the nucleotides in the editing window, they often convert other bases as well as the targeted one, which poses a difficulty and decreases the precision of base editing (Gaudelli *et al*., 2017, Komor *et al*., 2016).

Many different approaches to changing the width of the editing window were developed. Utilizing deaminases with decreased enzyme activity narrows the editing window, similarly, as shortening the sgRNA does, in some cases (Kim *et al*., 2017). On the other hand, tethering multiple deaminases to one nCas9 protein has been shown to enlarge it. This can be observed in the CBE-derived base editor for programming larger C to U (T) scope (BE-PLUS), which was engineered to effectively edit in the positions from 4 to 16 of protospacer (Jiang *et al*., 2018).

5.4. Prime editing

Prime editing is an extremely versatile strategy of gene editing able to induce small insertions or small deletions as well as all 12 base-to-base conversions (Anzalone *et al*., 2019).

Similarly as in base editing, this strategy does not require the creation of DSBs and, thus, the canonical Cas9 is not a part of the prime editing complex. Instead, it is altered by modifying its HNH nuclease domain, creating a Cas9n H840A mutant, which nicks only the non-complementary strand of the DNA. The whole system consists of the already mentioned Cas9n fused with reverse transcriptase and supplied with prime editing guide RNA (pegRNA), which contains the typical spacer but is also extended on its 3' end. This critical 3' end extension carries the template, according to which the reverse transcriptase incorporates the desired edit into the genome. Furthermore, a short sequence homologous with the 3' end of the DNA strand which is not complementary to the pegRNA is termed a primer binding site (PBS) and is located adjacent to this template (Anzalone *et al*., 2019).

In the beginning, the spacer contained in the pegRNA recognizes the target sequence and leads the Cas9n and the reverse transcriptase to the chosen genomic site. After the complex binds to the DNA, the R-loop is formed and Cas9n nicks the non-complementary strand. The PBS then anneals to the 3' end of the nicked strand and primes the reverse transcription of the template encoding the desired edit. Therefore, a 3' edited flap is created, which then must anneal to the complementary DNA strand for the edit to be installed. The annealing of the displaced 5' non-edited strand is thermodynamically favoured, however, endogenous nucleases have a preference for degrading such 5' flaps. Thus, the non-edited flap is excised, enabling the edited flap to anneal. Finally, the desired edit is copied to the complementary DNA strand by replication mechanism or repair pathways resulting in an altered DNA sequence (Anzalone *et al*., 2019).

This method of editing is very precise and known for inducing only little to no indels or off-target effects. The rarity of off-target effects is presumably caused by the requirement of three different complementary base pairing events. First, the spacer of pegRNA must hybridize with the target sequence to effectively lead the prime editor system to the chosen genomic site. Then, the PSB needs to pair with the 3' end of the nicked DNA strand to prime reverse transcription. Finally, the 3' flap is required to be partially homologous to the downstream DNA sequence for its efficient incorporation to happen. Nevertheless, prime editing has not yet reached the editing efficiency of base editing and other gene editing strategies associated with the CRISPR/Cas9 technology (Anzalone *et al*., 2019).

5.4.1. Prime editors

The first generation of prime editors (PE1) was created by fusing wild-type reverse transcriptase from Moloney murine leukaemia virus (M-MLV) with the C-terminus of *Sp*Cas9n and adding a specifically designed pegRNA. This system is capable of small insertions or deletions and single-base substitutions, however, its maximal editing efficiency is rather low (Anzalone *et al*., 2019).

To increase the rate of successfully edited sites, the second generation was developed by introducing various enhancing mutations to M-MLV reverse transcriptase. These mutations improved its processivity, thermostability and substrate affinity ultimately resulting in a 1.6- to 5.1-fold increase in the efficiency of editing (Anzalone *et al*., 2019).

By supplying the complex with another sgRNA, the third generation of prime editors (PE3) was established. The additional sgRNA directs the *Sp*Cas9n to nick the complementary strand about 40 to 90 nucleotides away from the first nick in the non-complementary strand. This causes the mismatch repair to preferably use the edited non-complementary strand as a repair template (Anzalone *et al*., 2019). This innovation improved efficiency by 1.5 to 4.2-fold over the previous generation, nevertheless, the increased frequency of indel formation was also reported. These indels presumably originated from the non-homologous repair of the DSB which was created by the simultaneous presence of single-stranded nicks in both DNA strands. To overcome this difficulty, the PE3b was developed. This system utilizes a sgRNA whose spacer is homologous to the reverse transcribed strand already containing the desired edit, not the original DNA sequence. As a result, the second nick is introduced only after the edited 3' flap resolution (Anzalone *et al*., 2019).

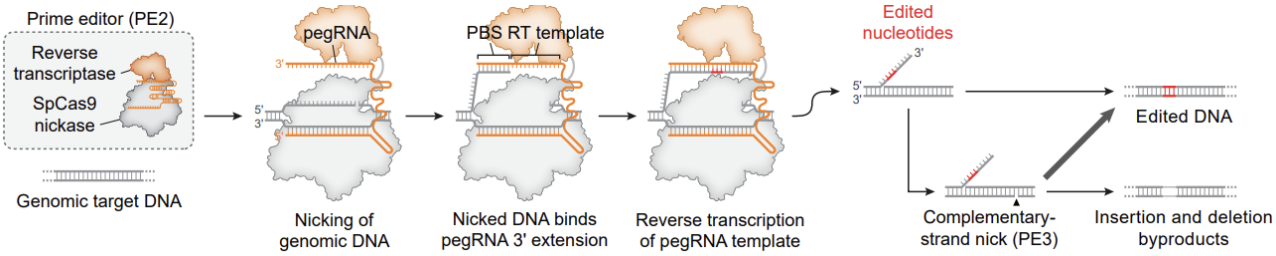

Figure 9: The schematic of prime editing
The second generation of the prime editors (PE2) consists of Cas9 nickase associated with reverse transcriptase and is guided to the target genomic site by pegRNA. The Cas9n nicks the edited strand and allows the complementary binding of the genomic strand and the PBS. The 3' extended end of pegRNA encodes the desired edit which is transferred into the genome by the synthetic activity of the reverse transcriptase. The edited strand is then incorporated into the genome. In case of the third generation of prime editors (PE3), the non-edited strand is nicked by additional nickase increasing the editing efficiency, and only after is the edit properly incorporated. (Adopted from Chen et al., 2021 and modified.)

To create a mouse model of cataract disorder, Lin J. *et al*. chose the third generation of prime editors and installed a specific single-nucleotide mutation into the mouse *Crygc* gene. Cataract disorder is characterized by the formation of cloudy area in the eye lens which leads to impaired vision.

It is caused by a deletion of a single G nucleotide in exon 3 of the *Crygc* gene which encodes the γC-crystallin protein. This mutation results in the creation of a downstream premature termination codon and, ultimately, in gene inactivation (Zhao *et al*., 2010). Lin J. and group used a PE3 system encoded by multiple plasmids and microinjected it directly into mouse embryos to install the desired edit in their genome. Using this strategy, they successfully generated mouse models displaying typical cataract phenotypes. In addition, they utilized the same PE3 system, only with different pegRNA, to repair this mutation in a cellular model by a single G insertion. This showed the potential use of the prime editing system in treating rare diseases (Lin J. *et al*., 2021).

Later, it was discovered that the DNA mismatch repair pathway causes the formation of undesired indels and generally counteracts the prime editing process. Therefore, MLH1dn protein, engineered to repress this repair mechanism, was expressed along with the original PE2 and PE3 systems creating PE4 and PE5, respectively. The fourth generation increased the editing efficiency by 7.7-fold over PE2, while the fifth generation was reported to enhance editing by 2-fold over PE3 in six of the tested cell types, including induced pluripotent stem cells (Chen *et al*., 2021).

As ineffective nuclear localization was proven to reduce the editing activity, Chen *et al*. altered the PE2 by modifying its original NLS and supplying the system with an additional one, resulting in the creation of PEmax. This optimized system also includes mutations in *Sp*Cas9, previously reported to enhance its activity, a 34 amino acids long linker connecting *Sp*Cas9n to RT, and an engineered reverse transcriptase domain bearing human codon-optimization. The same modifications can be introduced to other prime editor generations as well, considerably improving the efficiency of prime editing (Chen *et al*., 2021).

5.4.2. Alteration of pegRNA

The 3’ end extension of pegRNA is critical for successful prime editing, nevertheless, it is not protected from exonucleases. These enzymes can cause its degradation damaging the whole prime editor complex in the process, which results in inefficient editing. A way to overcome this difficulty is the addition of various structural motifs to the 3’ end of pegRNA to restrict the exonucleases from accessing it. Nelson *et al*. decided to add either a smaller modified prequeosine1-1 riboswitch aptamer (evopreQ1) or a larger pseudoknot from Moloney murine leukemia virus (mpknot) and connected it to the PBS via an 8-nt linker. This resulted in engineered pegRNAs (epegRNAs) with increased editing efficiency (Nelson *et al*., 2022). Another motif added to pegRNA to enhance its properties was, for example, an exoribonuclease resistant RNA motif extracted from Zikavirus. This structure was directly fused with the pegRNA without the need for a linker and displayed comparable editing efficiencies as previously mentioned epegRNAs (Zhang G. *et al*., 2022). Finally, Liu *et al*. fused pegRNA with a 20-nt Csy4 recognition motif hairpin. This hairpin protects the pegRNA from exonucleases and increases the editing efficiency. It also prevents the pegRNA

from circularization by restricting the PBS at its 3' end from self-annealing and binding to the complementary sequence of the spacer on its 5' end. (Liu *et al*., 2021).

The altered pegRNA played a role in the successful generation of a syndactyly mouse model by Qian and group using CRISPR/Cas9-mediated prime editing strategy. Syndactyly is a rare disease observed in humans caused by mutations in the *HOXD13* gene which leads to the fusion of fingers (Dai *et al*., 2014). The canonical PE3 was enhanced resulting in the creation of PE3max. This system was, subsequently, supplied with engineered pegRNA and the final ePE3max was generated. It was used to install a simple single-nucleotide G into T conversion into the mouse *Hoxd13* gene and, this way, a mouse model with an obvious finger fusion phenotype was introduced (Qian *et al*., 2023).

5.4.3. Prime editing using two pegRNAs

Using two pegRNAs, instead of just one, enables the prime editor system to introduce larger and more diverse changes into the genome. This way, both of the DNA strands can be edited at the same time, and many strategies taking advantage of this innovation were proposed in the scientific field.

First, Lin Q. *et al*. introduced the dual-pegRNA. This approach utilizes two pegRNAs, each targeting a specific site in one of the DNA strands in relative proximity to each other. After the Cas9n-mediated nicking of both strands, each reverse transcriptase synthesizes the 3' flap according to the template encoded in its associated pegRNA. These flaps are homologous to each other which facilitates their annealing and displacement of 5' flaps as well as their subsequent excision. This results in the successful incorporation of the edit into the target genomic site. Dual-pegRNA method creates single-base substitutions, small insertions and small deletions with higher editing efficiency than prime editors which use only one pegRNA. (Lin Q. *et al*., 2021). Not long after, Zhuang *et al*. developed the homologous 3′ extension mediated prime editor (HOPE) with similar properties to dual-pegRNAs and with the same range of possible gene modifications (Zhuang *et al*., 2022).

Another technique using paired peg-RNAs is PRIME-Del established by Choi *et al*. This strategy can introduce large deletions of up to 10 kbp while simultaneously inserting rather short sequences (approximately 30 bp) into the target site. Each of the two pegRNAs targets one of the complementary DNA strands which is then nicked by the associated Cas9n. The template coding the desired edit is then reverse transcribed creating 3' overhangs. These overhangs are complementary with each other and each of them also carries homology to the DNA sequence past the opposing nick. Finally, these edited 3' flaps align and the non-edited 5' flaps are displaced and excised. As a result, the DNA sequence between the two nicks is deleted and, potentially, replaced with a short exogenous insert (Choi *et al*., 2022).

The PE-Cas9-based deletion and repair (PEDAR) is a method of prime editing also employing two pegRNAs, which is able to create large deletions of up to 10 kbp and small insertions of approximately 60 bp. Similarly as in previous strategies, each of the two pegRNAs targets one of the DNA strands. This time, however, Cas9n is replaced with fully active Cas9 nuclease to introduce

a DSB into each target site. The template in the pegRNA is reverse transcribed creating two homologous 3' flaps in each strand, which subsequently anneal without the need for displacement and excision of the 5' unedited overhangs. Due to the previous formation of DSBs at the target sites, this intermediate is then processed by MMEJ or single-strand annealing (SSA) pathways, resulting in the incorporation of the desired edit. Unfortunately, the number of indel mutations exceeds the number of precise edits installed by this method (Jiang *et al*., 2022).

Finally, Anzalone *et al*. developed the twin-PE approach capable of introducing large deletions of up to 10 kbp and insertions of up to 113 bp into the chosen genomic site. Each of the pegRNAs, again, targets one of the opposing DNA strands and templates the reverse transcription of the desired edit. The 3' edited flaps are complementary with each other, nevertheless, they carry no homology to the DNA past the opposing nick. This modification further enhances their preference of annealing to one another and, thus, the editing efficiency of the whole system as well. After the 5' non-edited overhangs are excised, the intended alteration is incorporated into the genome, similarly as in previous methods. This technique proved to be capable of installing accurate edits with very high efficiency (Anzalone *et al*., 2022).

6. Conclusion

Since CRISPR/Cas9 was discovered and modified in a way that allowed it to be effectively used in research, this technique became a popular choice for gene editing experiments. Its low cost, simplicity and wide repertoire of possible edits are the main benefits which convinced scientists to adopt this system. Today, numerous different gene editing strategies taking advantage of the CRISPR/Cas9 system exist, each specifically designed to successfully install the intended edit and bearing its own advantages and disadvantages. The most well-known approaches include, for instance, gene knock-out caused by Cas9-mediated introduction of DSB which is usually repaired by the imprecise NHEJ pathway. This repair results in random indels which can lead to frameshift mutations and the creation of premature STOP codon, degradation of the transcript and, ultimately, the gene inactivation (Couttet and Grange, 2004, Yen *et al*., 2014). Gene knock-out can be also achieved by the deletion of the specific critical genomic sequence using a dual gRNA strategy (Canver *et al*., 2014). Another frequently used method is gene knock-in in which a specific sequence, for example synthetic fragment of DNA or a transgene, is incorporated into the genome and, subsequently, expressed (Raveux *et al*., 2017). Finally, base editing enables researchers to convert specific nucleotides into others while prime editing is capable of all 12 base substitutions as well as installing small insertions and deletions (Anzalone *et al*., 2019, Komor *et al*., 2016).

Even though CRISPR/Cas9 has numerous undeniable advantages, many further modifications and upgrades have to be implemented to the system to achieve its maximum potential. Some of the main challenges researchers face when using this technique include PAM restrictions which can make

targeting the genomic site of interest rather difficult, in some cases (Kim *et al*., 2017). Off-target effects are a result of installing the edits into non-targeted genomic sites and pose another important difficulty. These non-targeted sites are only partially homologous to the sgRNA, nevertheless, they are still able to pair with sgRNA due to the low number of mismatched bases being tolerated (Fu *et al*., 2013). Many modifications addressing these issues have already been successfully applied (Kim *et al*., 2017, Ran *et al*., 2013), however, there is still a need for additional research.

Considering all the qualities of CRISPR/Cas9 mentioned above, it is no surprise that this method was adopted for the creation of mouse models representing human rare diseases. Many mice characteristics, such as small size, low cost, short lifespan and generation time and, finally, the similarity of their genome to the human one, predetermine them to be perfect experimental systems for modelling human disorders (Mouse Genome Sequencing Consortium, 2002, Vandamme, 2015). Using mouse models to help scientists better understand the causalities and molecular mechanisms of rare diseases was proven to be an efficient and relatively easy strategy. In addition, CRISPR/Cas9 has been shown to create precise edits and generate mouse models mimicking rare diseases in humans rather accurately (Borrás *et al*., 2020, Duan *et al*., 2016, Liang *et al*., 2017, Liang *et al*., 2018, Lin J. *et al*., 2021, Nakagawa *et al*., 2016, Nutter *et al*., 2019, Qian *et al*., 2023, Rauch *et al*., 2018, Syding *et al*., 2022, Zhang Y. *et al*., 2022). This approach holds a promise for people suffering from such conditions as it enables scientists to conduct research more effectively. Data derived from such experiments should enhance our knowledge of rare diseases and make the necessary information as well as novel treatment options accessible to the general public.

Nevertheless, there still are some challenges ultimately decreasing the accuracy of mice models. Despite the similarities in their genetic information, only about 40 % of the human genome can be aligned to the mouse one, which is no surprise considering their obvious anatomical and physiological differences (Mouse Genome Sequencing Consortium, 2002). Moreover, the majority of research is conducted on inbred mouse strains to reduce results variability and ensure the reproducibility of the experiments. However, it has been shown that a single gene alteration causes dramatically different clinical manifestations in mice from various genetic backgrounds. This suggests that some experimental results are not easily translatable even to other mouse strains, let alone to humans (Sittig *et al*., 2016). Finally, experimenting on animals is quite controversial despite its undeniable benefits. The use of animal models must follow strict rules and regulations and be carefully evaluated as well. Furthermore, only experienced personnel are allowed to work with such models for animal welfare to be ensured. Complete elimination of experimenting on animals is, unfortunately, not an option as mice remain the most convenient modelling system for numerous branches of scientific research (Fenwick *et al*., 2009).

7. List of literature

Secondary sources are marked with *

Aida T, Chiyo K, Usami T, Ishikubo H, Imahashi R, Wada Y, Tanaka KF, Sakuma T, Yamamoto T, Tanaka K (2015). Cloning-free CRISPR/Cas system facilitates functional cassette knock-in in mice. Genome Biol. *16*, 87.

Aida T, Nakade S, Sakuma T, Izu Y, Oishi A, Mochida K, Ishikubo H, Usami T, Aizawa H, Yamamoto T, Tanaka K (2016). Gene cassette knock-in in mammalian cells and zygotes by enhanced MMEJ. BMC Genomics *17*, 979.

Akagi K, Sandig V, Vooijs M, Van der Valk M, Giovannini M, Strauss M, Berns A (1997). Cre-mediated somatic site-specific recombination in mice. Nucleic Acids Res. *25*, 1766-1773.

Anzalone AV, Gao XD, Podracky CJ, Nelson AT, Koblan LW, Raguram A, Levy JM, Mercer JAM, Liu DR (2022). Programmable deletion, replacement, integration and inversion of large DNA sequences with twin prime editing. Nat. Biotechnol. *40*, 731-740.

Anzalone AV, Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, Chen PJ, Wilson C, Newby GA, Raguram A, Liu DR (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. Nature *576*, 149-157.

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

Beumer KJ, Trautman JK, Bozas A, Liu JL, Rutter J, Gall JG, Carroll D (2008). Efficient gene targeting in Drosophila by direct embryo injection with zinc-finger nucleases. Proc. Natl. Acad. Sci. U.S.A. *105*, 19821-19826.

Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, Bonas U (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. Science *326*, 1509-1512.

Borrás T, Cowley DO, Asokan P, Pandya K (2020). Generation of a Matrix Gla (Mgp) floxed mouse, followed by conditional knockout, uncovers a new Mgp function in the eye. Sci. Rep. *10*, 18583.

Bradley A, Anastassiadis K, Ayadi A, Battey JF, Bell C, Birling MC, Bottomley J, Brown SD, Bürger A, Bult CJ, Bushell W, Collins FS, Desaintes C, Doe B, Economides A, Eppig JT, Finnell RH, Fletcher C, Fray M, Frendewey D, Friedel RH, Grosveld FG, Hansen J, Hérault Y, Hicks G, Hörlein A, Houghton R, Hrabé de Angelis M, Huylebroeck D, Iyer V, de Jong PJ, Kadin JA, Kaloff C, Kennedy K, Koutsourakis M, Lloyd KC, Marschall S, Mason J, McKerlie C, McLeod MP, von Melchner H, Moore M, Mujica AO, Nagy A, Nefedov M, Nutter LM, Pavlovic G, Peterson JL, Pollock J, Ramirez-Solis R, Rancourt DE, Raspa M, Remacle JE, Ringwald M, Rosen B, Rosenthal N, Rossant J, Ruiz Noppinger P, Ryder E, Schick JZ, Schnütgen F, Schofield P, Seisenberger C, Selloum M, Simpson EM, Skarnes WC, Smedley D, Stanford WL, Stewart AF, Stone K, Swan K, Tadepally H, Teboul L, Tocchini-Valentini GP, Valenzuela D, West AP, Yamamura K, Yoshinaga Y, Wurst W. The mammalian gene function resource: the International Knockout Mouse Consortium (2012). Mamm. Genome *23*, 580-586.

Brook JD, McCurrach ME, Harley HG, Buckler AJ, Church D, Aburatani H, Hunter K, Stanton VP, Thirion JP, Hudson T, *et al* (1992). Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. Cell *68*, 799-808.

Brooks, P. J., Danilo A. Tagle, and Steve Groft (2014). Expanding Trials of Novel Therapeutics for Rare Diseases Based on Grouping by Molecular Etiology. Nat. Biotechnol. *32*, 515.

Brown SD, Moore MW (2012). Towards an encyclopaedia of mammalian gene function: the International Mouse Phenotyping Consortium. Dis. Model Mech. *5*, 289-292.

Byrne SM, Ortiz L, Mali P, Aach J, Church GM (2015). Multi-kilobase homozygous targeted gene replacement in human induced pluripotent stem cells. Nucleic Acids Res. *43*, e21-e21.

Canver MC, Bauer DE, Dass A, Yien YY, Chung J, Masuda T, Maeda T, Paw BH, Orkin SH (2014). Characterization of genomic deletion efficiency mediated by clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells. J. Biol. Chem. *289*, 21312-21324.

Chen L, Ye Y, Dai H, Zhang H, Zhang X, Wu Q, Zhu Z, Spalinskas R, Ren W, Zhang W (2018). User-Friendly Genetic Conditional Knockout Strategies by CRISPR/Cas9. Stem Cells Int. *2018*, 1-10.

Chen PJ, Hussmann JA, Yan J, Knipping F, Ravisankar P, Chen PF, Chen C, Nelson JW, Newby GA, Sahin M, Osborn MJ, Weissman JS, Adamson B, Liu DR (2021). Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. Cell *184*, 5635-5652.

Choi J, Chen W, Suiter CC, Lee C, Chardon FM, Yang W, Leith A, Daza RM, Martin B, Shendure J (2022). Precise genomic deletions using paired prime editing. Nat. Biotechnol. *40*, 218-226.

Chu VT, Weber T, Graf R, Sommermann T, Petsch K, Sack U, Volchkov P, Rajewsky K, Kühn R (2016). Efficient generation of Rosa26 knock-in mice using CRISPR/Cas9 in C57BL/6 zygotes. BMC Biotechnol. *16*, 1-15.

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

Couttet P, Grange T (2004). Premature termination codons enhance mRNA decapping in human cells. Nucleic Acids Res. *32*, 488-494.

d'Adda di Fagagna F, Weller GR, Doherty AJ, Jackson SP (2003). The Gam protein of bacteriophage Mu is an orthologue of eukaryotic Ku. EMBO Rep. *4*, 47-52.

Dai, L., Liu, D., Song, M., Xu, X., Xiong, G., Yang, K. & Bai, Y. (2014). Mutations in the homeodomain of HOXD13 cause syndactyly type 1-c in two Chinese families. PLoS One *9*, e96192.

Dianov G, Price A, Lindahl T (1992). Generation of single-nucleotide repair patches following excision of uracil residues from DNA. Mol. Cell Biol. *12*, 1605-1612.

Duan X, Liu J, Zheng X, Wang Z, Zhang Y, Hao Y, Yang T, Deng H (2016). Deficiency of ATP6V1H Causes Bone Loss by Inhibiting Bone Resorption and Bone Formation through the TGF-β1 Pathway. Theranostics *6*, 2183-2195

Fang H, Bygrave AM, Roth RH, Johnson RC, Huganir RL (2021). An optimized CRISPR/Cas9 approach for precise genome editing in neurons. Elife *10*, e65202.

Feil R, Brocard J, Mascrez B, LeMeur M, Metzger D, Chambon P (1996). Ligand-activated site-specific recombination in mice. Proc. Natl. Acad. Sci. U.S.A. *93*, 10887-10890.

Fenwick N, Griffin G, Gauthier C (2009). The welfare of animals used in science: how the "Three Rs" ethic guides improvements. Can. Vet. J. *50*, 523-530.

Fermaglich LJ, Miller KL (2023). A comprehensive study of the rare diseases and conditions targeted by orphan drug designations and approvals over the forty years of the Orphan Drug Act. Orphanet J. Rare Dis. *18*, 163.

Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, Sander JD (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat. Biotechnol. *31*, 822-826.

Fu YW, Dai XY, Wang WT, Yang ZX, Zhao JJ, Zhang JP, Wen W, Zhang F, Oberg KC, Zhang L, Cheng T, Zhang XB (2021). Dynamics and competition of CRISPR-Cas9 ribonucleoproteins and AAV donor-mediated NHEJ, MMEJ and HDR editing. Nucleic Acids Res. *49*, 969-985.

Garattini S, Grignaschi G (2017). Animal testing is still the best way to find new treatments for patients. Eur. J. Intern. Med. *39*, 32-35.

Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, Bryson DI, Liu DR (2017). Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. Nature *551*, 464-471.

Geisinger JM, Turan S, Hernandez S, Spector LP, Calos MP (2016). In vivo blunt-end cloning through CRISPR/Cas9-facilitated non-homologous end-joining. Nucleic Acids Res. *44*, e76-e76.

Gossen, M., Freundlieb, S., Bender, G., Müller, G., Hillen, W., & Bujard, H. (1995). Transcriptional activation by tetracyclines in mammalian cells. Science, *268*, 1766-1769.

Hall GW, Thein S (1994). Nonsense codon mutations in the terminal exon of the beta-globin gene are not associated with a reduction in beta-mRNA accumulation: a mechanism for the phenotype of dominant beta-thalassemia. Blood *83*, 2031-2037.

Han J, Zhang J, Chen L, Shen B, Zhou J, Hu B, Du Y, Tate PH, Huang X, Zhang W (2014). Efficient in vivo deletion of a large imprinted lncRNA by CRISPR/Cas9. RNA Biol. *11*, 829-835.

He X, Tan C, Wang F, Wang Y, Zhou R, Cui D, You W, Zhao H, Ren J, Feng B (2016). Knock-in of large reporter genes in human cells via CRISPR/Cas9-induced homology-dependent and independent DNA repair. Nucleic Acids Res. *44*, e85-e85.

Heeb LV, Taskoparan B, Katsoulas A, Beffinger M, Clavien PA, Kobold S, Gupta A, Berg JV (2023). HDR-based CRISPR/Cas9-mediated Knockout of PD-L1 in C57BL/6 Mice. Bio. Protoc. *13*, e4724.

Horii T, Morita S, Kimura M, Terawaki N, Shibutani M, Hatada I (2017). Efficient generation of conditional knockout mice via sequential introduction of lox sites. Sci. Rep. *7*, 7891.

Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O, Cradick TJ, Marraffini LA, Bao G, Zhang F (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nat. Biotechnol. *31*, 827-32.

Jansen R, Embden JD, Gaastra W, Schouls LM (2002). Identification of genes that are associated with DNA repeats in prokaryotes. Mol. Microbiol. *43*, 1565-1575.

Jiang T, Zhang XO, Weng Z, Xue W (2022). Deletion and replacement of long genomic sequences using prime editing. Nat. Biotechnol. *40*, 227-234.

Jiang W, Feng S, Huang S, Yu W, Li G, Yang G, Liu Y, Zhang Y, Zhang L, Hou Y, Chen J, Chen J, Huang X (2018). BE-PLUS: a new base editing tool with broadened editing window and enhanced fidelity. Cell Res. *28*, 855-861.

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

Kapahnke M, Banning A, Tikkanen R (2016). Random Splicing of Several Exons Caused by a Single Base Change in the Target Exon of CRISPR/Cas9 Mediated Gene Knockout. Cells *5*, 45.

Kim J, Malashkevich V, Roday S, Lisbin M, Schramm VL, Almo SC (2006). Structural and kinetic characterization of Escherichia coli TadA, the wobble-specific tRNA deaminase. Biochemistry *45*, 6407-6416.

Koblan LW, Doman JL, Wilson C, Levy JM, Tay T, Newby GA, Maianti JP, Raguram A, Liu DR (2018). Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. Nat. Biotechnol. *36*, 843-846.

Koenig M, Hoffman EP, Bertelson CJ, Monaco AP, Feener C, Kunkel LM (1987). Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. Cell *50*, 509-517.

Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature *533*, 420-424.

Komor AC, Zhao KT, Packer MS, Gaudelli NM, Waterbury AL, Koblan LW, Kim YB, Badran AH, Liu DR (2017). Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. Sci. Adv. *3*, eaao4774.

Körner A, Pawelek J (1982). Mammalian tyrosinase catalyzes three reactions in the biosynthesis of melanin. Science *217*, 1163-1165.

Kurt IC, Zhou R, Iyer S, Garcia SP, Miller BR, Langner LM, Grünewald J, Joung JK (2021). CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. Nat. Biotechnol. *39*, 41-46.

Lau AY, Wyatt MD, Glassner BJ, Samson LD, Ellenberger T (2000). Molecular basis for discriminating between normal and damaged bases by the human alkyladenine glycosylase, AAG. Proc. Natl. Acad. Sci. U.S.A. *97*, 13573-13578.

Leidy-Davis T, Cheng K, Goodwin LO, Morgan JL, Juan WC, Roca X, Ong ST, Bergstrom DE (2018). Viable Mice with Extensive Gene Humanization (25-kbp) Created Using Embryonic Stem Cell/Blastocyst and CRISPR/Zygote Injection Approaches. Sci. Rep. *8*, 15028.

Liang P, Sun H, Zhang X, Xie X, Zhang J, Bai Y, Ouyang X, Zhi S, Xiong Y, Ma W, Liu D, Huang J, Songyang Z (2018). Effective and precise adenine base editing in mouse zygotes. Protein Cell *9*, 808-813.

Liang P, Sun H, Sun Y, Zhang X, Xie X, Zhang J, Zhang Z, Chen Y, Ding C, Xiong Y, Ma W, Liu D, Huang J, Songyang Z (2017). Effective gene editing by high-fidelity base editor 2 in mouse zygotes. Protein Cell *8*, 601-611.

Lin J, Liu X, Lu Z, Huang S, Wu S, Yu W, Liu Y, Zheng X, Huang X, Sun Q, Qiao Y, Liu Z (2021). Modeling a cataract disorder in mice with prime editing. Mol. Ther. Nucleic Acids *25*, 494-501.

Lin Q, Jin S, Zong Y, Yu H, Zhu Z, Liu G, Kou L, Wang Y, Qiu JL, Li J, Gao C (2021). High-efficiency prime editing with optimized, paired pegRNAs in plants. Nat. Biotechnol. *39*, 923-927.

Lindahl T (1974). An N-glycosidase from Escherichia coli that releases free uracil from DNA containing deaminated cytosine residues. Proc. Natl. Acad. Sci. U.S.A. *71*, 3649-3653.

Lindhurst MJ, Fiermonte G, Song S, Struys E, De Leonardis F, Schwartzberg PL, Chen A, Castegna A, Verhoeven N, Mathews CK, Palmieri F, Biesecker LG (2006). Knockout of Slc25a19 causes mitochondrial thiamine pyrophosphate depletion, embryonic lethality, CNS malformations, and anemia. Proc. Natl. Acad. Sci. U.S.A. *103*, 15927-15932.

Liu Y, Yang G, Huang S, Li X, Wang X, Li G, Chi T, Chen Y, Huang X, Wang X (2021). Enhancing prime editing by Csy4-mediated processing of pegRNA. Cell Res. *31*, 1134-1136.

Luo G, Ducy P, McKee MD, Pinero GJ, Loyer E, Behringer RR, Karsenty G (1997). Spontaneous calcification of arteries and cartilage in mice lacking matrix GLA protein. Nature *386*, 78-81.

Mani M, Smith J, Kandavelou K, Berg JM, Chandrasegaran S (2005). Binding of two zinc finger nuclease monomers to two specific sites is required for effective double-strand DNA cleavage. Biochem. Biophys. Res. Commun. *334*, 1191-1197.

Miura H, Quadros RM, Gurumurthy CB, Ohtsuka M (2018). Easi-CRISPR for creating knock-in and conditional knockout mouse models using long ssDNA donors. Nat. Protoc. *13*, 195-215.

Mojica FJM, Díez-Villaseñor C, García-Martínez J, Almendros C (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology *155*, 733-740.

Mouse Genome Sequencing Consortium; Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, Antonarakis SE, Attwood J, Baertsch R, Bailey J, Barlow K, Beck S, Berry E, Birren B, Bloom T, Bork P, Botcherby M, Bray N, Brent MR, Brown DG, Brown SD, Bult C, Burton J, Butler J, Campbell RD, Carninci P, Cawley S, Chiaromonte F, Chinwalla AT, Church DM, Clamp M, Clee C, Collins FS, Cook LL, Copley RR, Coulson A, Couronne O, Cuff J, Curwen V, Cutts T, Daly M, David R, Davies J, Delehaunty KD, Deri J, Dermitzakis ET, Dewey C, Dickens NJ, Diekhans M, Dodge S, Dubchak I, Dunn DM, Eddy SR, Elnitski L, Emes RD, Eswara P, Eyras E, Felsenfeld A, Fewell GA, Flicek P, Foley K, Frankel WN, Fulton LA, Fulton RS, Furey TS, Gage D, Gibbs RA, Glusman G, Gnerre S, Goldman N, Goodstadt L, Grafham D, Graves TA, Green ED, Gregory S, Guigó R, Guyer M, Hardison RC, Haussler D, Hayashizaki Y, Hillier LW, Hinrichs A, Hlavina W, Holzer T, Hsu F, Hua A, Hubbard T, Hunt A, Jackson I, Jaffe DB, Johnson LS, Jones M, Jones TA, Joy A, Kamal M, Karlsson EK, Karolchik D, Kasprzyk A, Kawai J, Keibler E, Kells C, Kent WJ, Kirby A, Kolbe DL, Korf I, Kucherlapati RS, Kulbokas EJ, Kulp D, Landers T, Leger JP, Leonard S, Letunic I, Levine R, Li J, Li M, Lloyd C, Lucas S, Ma B, Maglott DR, Mardis ER, Matthews L, Mauceli E, Mayer JH, McCarthy M, McCombie WR, McLaren S, McLay K, McPherson JD, Meldrim J, Meredith B, Mesirov JP, Miller W, Miner TL, Mongin E, Montgomery KT, Morgan M, Mott R, Mullikin JC, Muzny DM, Nash WE, Nelson JO, Nhan MN, Nicol R, Ning Z, Nusbaum C, O'Connor MJ, Okazaki Y, Oliver K, Overton-Larty E, Pachter L, Parra G, Pepin KH, Peterson J, Pevzner P, Plumb R, Pohl CS, Poliakov A, Ponce TC, Ponting CP, Potter S, Quail M, Reymond A, Roe BA, Roskin KM, Rubin EM, Rust AG, Santos R, Sapojnikov V, Schultz B, Schultz J, Schwartz MS, Schwartz S, Scott C, Seaman S, Searle S, Sharpe T, Sheridan A, Shownkeen R, Sims S, Singer JB, Slater G, Smit A, Smith DR, Spencer B, Stabenau A, Stange-Thomann N, Sugnet C, Suyama M, Tesler G, Thompson J, Torrents D, Trevaskis E, Tromp J, Ucla C, Ureta-Vidal A, Vinson JP, Von Niederhausern AC, Wade CM, Wall M, Weber RJ, Weiss RB, Wendl MC, West AP, Wetterstrand K, Wheeler R, Whelan S, Wierzbowski J, Willey D, Williams S, Wilson RK, Winter E, Worley KC, Wyman D, Yang S, Yang SP, Zdobnov EM, Zody MC, Lander ES (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520-562.

Munroe PB, Olgunturk RO, Fryns JP, Van Maldergem L, Ziereisen F, Yuksel B, Gardiner RM, Chung E (1999). Mutations in the gene encoding the human matrix Gla protein cause Keutel syndrome. Nat. Genet. *21*,142-144.

Nakade S, Tsubota T, Sakane Y, Kume S, Sakamoto N, Obara M, Daimon T, Sezutsu H, Yamamoto T, Sakuma T, Suzuki KT (2014). Microhomology-mediated end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9. Nat. Commun. *5*, 5560.

Nakagawa Y, Oikawa F, Mizuno S, Ohno H, Yagishita Y, Satoh A, Osaki Y, Takei K, Kikuchi T, Han SI, Matsuzaka T, Iwasaki H, Kobayashi K, Yatoh S, Yahagi N, Isaka M, Suzuki H, Sone H, Takahashi S, Yamada N, Shimano H (2016). Hyperlipidemia and hepatitis in liver-specific CREB3L3 knockout mice generated using a one-step CRISPR/Cas9 system. Sci. Rep. *6*, 27857.

Nelson JW, Randolph PB, Shen SP, Everette KA, Chen PJ, Anzalone AV, An M, Newby GA, Chen JC, Hsu A, Liu DR (2022). Engineered pegRNAs improve prime editing efficiency. Nat. Biotechnol. *40*, 402-410.

Neu-Yilik G, Amthor B, Gehring NH, Bahri S, Paidassi H, Hentze MW, Kulozik AE (2011). Mechanism of escape from nonsense-mediated mRNA decay of human beta-globin transcripts with nonsense mutations in the first exon. RNA *17*, 843-854.

Nguengang Wakap S, Lambert DM, Olry A, Rodwell C, Gueydan C, Lanneau V, Murphy D, Le Cam Y, Rath A. (2020). Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. Eur. J. Hum. Genet. *28*, 165-173.

Nishida K, Arazoe T, Yachie N, Banno S, Kakimoto M, Tabata M, Mochizuki M, Miyabe A, Araki M, Hara KY, Shimatani Z, Kondo A (2016). Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. Science *353*, aaf8729.

Nutter CA, Bubenik JL, Oliveira R, Ivankovic F, Sznajder ŁJ, Kidd BM, Pinto BS, Otero BA, Carter HA, Vitriol EA, Wang ET, Swanson MS (2019). Cell-type-specific dysregulation of RNA alternative splicing in short tandem repeat mouse knockin models of myotonic dystrophy. Genes Dev. *33*, 1635-1640.

Orphan Drug Act, Pub L. No. 97–414, 96 Stat. 2049 (1983).

Pettitt SJ, Liang Q, Rairdan XY, Moran JL, Prosser HM, Beier DR, Lloyd KC, Bradley A, Skarnes WC (2009). Agouti C57BL/6N embryonic stem cells for mouse genetic resources. Nat. Methods *6*, 493-495.

Qian Y, Wang D, Niu W, Shi Z, Wu M, Zhao D, Li J, Gao X, Zhang Z, Lai L, Li Z (2023). Development of a highly efficient prime editor system in mice and rabbits. Cell. Mol. Life Sci. *80*, 346.

Quadros RM, Miura H, Harms DW, Akatsuka H, Sato T, Aida T, Redder R, Richardson GP, Inagaki Y, Sakai D, Buckley SM, Seshacharyulu P, Batra SK, Behlke MA, Zeiner SA, Jacobi AM, Izu Y, Thoreson WB, Urness LD, Mansour SL, Ohtsuka M, Gurumurthy CB (2017). Easi-CRISPR: a robust method for one-step generation of mice carrying conditional and insertion alleles using long ssDNA donors and CRISPR ribonucleoproteins. Genome Biol. *18*, 92.

Ran FA, Hsu PD, Lin CY, Gootenberg JS, Konermann S, Trevino AE, Scott DA, Inoue A, Matoba S, Zhang Y, Zhang F (2013). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell *154*, 1380-1389.

Rare Disease Act of 2002, Pub L. No. 107–280, 116 Stat. 1988 (2002).

Rath A, Olry A, Dhombres F, Brandt MM, Urbero B, Ayme S (2012). Representation of rare diseases in health information systems: the Orphanet approach to serve a wide range of end users. Hum. Mutat. *33*, 803-8.

Rauch F, Geng Y, Lamplugh L, Hekmatnejad B, Gaumond MH, Penney J, Yamanaka Y, Moffatt P (2018). Crispr-Cas9 engineered osteogenesis imperfecta type V leads to severe skeletal deformities and perinatal lethality in mice. Bone *107*, 131-142.

Raveux A, Vandormael-Pournin S, Cohen-Tannoudji M (2017). Optimization of the production of knock-in alleles by CRISPR/Cas9 microinjection into the mouse zygote. Sci. Rep. *7*, 42661.

Reber S, Mechtersheimer J, Nasif S, Benitez JA, Colombo M, Domanski M, Jutzi D, Hedlund E, Ruepp MD (2018). CRISPR-Trap: a clean approach for the generation of gene knockouts and gene replacements in human cells. Mol. Biol. Cell *29*, 75-83.

Regulation (EC) No. 141/2000 of the European Parliament and of the Council of 16 December 1999 on orphan medicinal products (2000) OJ L 18/1.

Richter MF, Zhao KT, Eton E, Lapinaite A, Newby GA, Thuronyi BW, Wilson C, Koblan LW, Zeng J, Bauer DE, Doudna JA, Liu DR (2020). Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. Nat. Biotechnol. *38*, 883-891.

*Richter T, Nestler-Parr S, Babela R, Khan ZM, Tesoro T, Molsen E, Hughes DA; International Society for Pharmacoeconomics and Outcomes Research Rare Disease Special Interest Group (2015). Rare Disease Terminology and Definitions-A Systematic Global Review: Report of the ISPOR Rare Disease Special Interest Group. Value Health *18*, 906-14.

Sakuma T, Nakade S, Sakane Y, Suzuki KT, Yamamoto T (2016). MMEJ-assisted gene knock-in using TALENs and CRISPR-Cas9 with the PITCh systems. Nat. Protoc. *11*, 118-133.

Semler O, Garbes L, Keupp K, Swan D, Zimmermann K, Becker J, Iden S, Wirth B, Eysel P, Koerber F, Schoenau E, Bohlander SK, Wollnik B, Netzer C (2012). A mutation in the 5'-UTR of IFITM5 creates an in-frame start codon and causes autosomal-dominant osteogenesis imperfecta type V with hyperplastic callus. Am. J. Hum. Genet. *91*, 349-357.

Sittig LJ, Carbonetto P, Engel KA, Krauss KS, Barrios-Camacho CM, Palmer AA (2016). Genetic Background Limits Generalizability of Genotype-Phenotype Relationships. Neuron *91*, 1253-1259.

Skarnes WC, Rosen B, West AP, Koutsourakis M, Bushell W, Iyer V, Mujica AO, Thomas M, Harrow J, Cox T, Jackson D, Severin J, Biggs P, Fu J, Nefedov M, de Jong PJ, Stewart AF, Bradley A (2011). A conditional knockout resource for the genome-wide study of mouse gene function. Nature *474*, 337-342.

Song F, Stieger K (2017). Optimizing the DNA Donor Template for Homology-Directed Repair of Double-Strand Breaks. Mol. Ther. Nucleic Acids *7*, 53-60.

Soriano P (1999). Generalized lacZ expression with the ROSA26 Cre reporter strain. Nat. Genet. *21*, 70-71.

Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature *507*, 62-67.

Suzuki K, Tsunekawa Y, Hernandez-Benitez R, Wu J, Zhu J, Kim EJ, Hatanaka F, Yamamoto M, Araoka T, Li Z, Kurita M, Hishida T, Li M, Aizawa E, Guo S, Chen S, Goebl A, Soligalla RD, Qu J, Jiang T, Fu X, Jafari M, Esteban CR, Berggren WT, Lajara J, Nuñez-Delicado E, Guillen P, Campistol JM, Matsuzaki F, Liu GH, Magistretti P, Zhang K, Callaway EM, Zhang K, Belmonte JC (2016). In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. Nature *540*, 144-149.

Syding LA, Kubik-Zahorodna A, Nickl P, Novosadova V, Kopkanova J, Kasparek P, Prochazka J, Sedlacek R (2022). Generation and Characterization of a Novel Angelman Syndrome Mouse Model with a Full Deletion of the Ube3a Gene. Cells *11*, 2815.

Szczelkun MD, Tikhomirova MS, Sinkunas T, Gasiunas G, Karvelis T, Pschera P, Siksnys V, Seidel R (2014). Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. Proc. Natl. Acad. Sci. U.S.A. *111*, 9798-9803.

Takata M, Sasaki MS, Sonoda E, Morrison C, Hashimoto M, Utsumi H, Yamaguchi-Iwai Y, Shinohara A, Takeda S (1998). Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. EMBO J. *17*, 5497-5508.

Testa G, Schaft J, van der Hoeven F, Glaser S, Anastassiadis K, Zhang Y, Hermann T, Stremmel W, Stewart AF (2004). A reliable lacZ expression reporter cassette for multipurpose, knockout-first alleles. Genesis *38*, 151-158.

*Vandamme TF (2015). Rodent models for human diseases. Eur. J. Pharmacol. *759*, 84-89.

Wang H, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, Zhang F, Jaenisch R (2013). One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell *153*, 910-918.

Wu SS, Lee H, Szép-Bakonyi R, Colozza G, Boese A, Gert KR, Hallay N, Lee JH, Kim J, Zhu Y, Linssen MM, Pilat-Carotta S, Hohenstein P, Theussl HC, Pauli A, Koo BK (2022). SCON-a Short Conditional intrON for conditional knockout with one-step zygote injection. Exp. Mol. Med. *54*, 2188-2199.

Wu Y, Liang D, Wang Y, Bai M, Tang W, Bao S, Yan Z, Li D, Li J (2013). Correction of a genetic disease in mouse via use of CRISPR-Cas9. Cell Stem Cell *13*, 659-662.

Xie J, Huang X, Wang X, Gou S, Liang Y, Chen F, Li N, Ouyang Z, Zhang Q, Ge W, Jin Q, Shi H, Zhuang Z, Zhao X, Lian M, Wang J, Ye Y, Quan L, Wu H, Wang K, Lai L (2020). ACBE, a new base editor for simultaneous C-to-T and A-to-G substitutions in mammalian systems. BMC Biol. *18*, 131.

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

Yao X, Zhang M, Wang X, Ying W, Hu X, Dai P, Meng F, Shi L, Sun Y, Yao N, Zhong W, Li Y, Wu K, Li W, Chen ZJ, Yang H (2018). Tild-CRISPR Allows for Efficient and Precise Gene Knockin in Mouse and Human Cells. Dev Cell *45*, 526-536.

Yeh WH, Chiang H, Rees HA, Edge ASB, Liu DR (2018). In vivo base editing of post-mitotic sensory cells. Nat. Commun. *9*, 2184.

Yen ST, Zhang M, Deng JM, Usman SJ, Smith CN, Parker-Thornburg J, Swinton PG, Martin JF, Behringer RR (2014). Somatic mosaicism and allele complexity induced by CRISPR/Cas9 RNA injections in mouse zygotes. Dev. Biol. *393*, 3-9.

Zambrowicz BP, Imamoto A, Fiering S, Herzenberg LA, Kerr WG, Soriano P (1997). Disruption of overlapping transcripts in the ROSA beta geo 26 gene trap strain leads to widespread expression of beta-galactosidase in mouse embryos and hematopoietic cells. Proc. Natl. Acad. Sci. U.S.A. *94*, 3789-3794.

Zare K, Shademan M, Ghahramani Seno MM, Dehghani H (2018). CRISPR/Cas9 Knockout Strategies to Ablate CCAT1 lncRNA Gene in Cancer Cells. Biol. Proced. Online *20*, 1-12.

Zhang G, Liu Y, Huang S, Qu S, Cheng D, Yao Y, Ji Q, Wang X, Huang X, Liu J (2022). Enhancement of prime editing via xrRNA motif-joined pegRNA. Nat. Commun. *13*, 1856.

Zhang J, Maquat LE (1997). Evidence that translation reinitiation abrogates nonsense-mediated mRNA decay in mammalian cells. EMBO J. *16*, 826-833.

Zhang JP, Li XL, Li GH, Chen W, Arakaki C, Botimer GD, Baylink D, Zhang L, Wen W, Fu YW, Xu J, Chun N, Yuan W, Cheng T, Zhang XB (2017). Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage. Genome Biol. *18*, 35.

Zhang Y, Li H, Nishiyama T, McAnally JR, Sanchez-Ortiz E, Huang J, Mammen PPA, Bassel-Duby R, Olson EN (2022). A humanized knockin mouse model of Duchenne muscular dystrophy and its correction by CRISPR-Cas9 therapeutic gene editing. Mol. Ther. Nucleic Acids *29*, 525-537.

Zhao L, Li K, Bao S, Zhou Y, Liang Y, Zhao G, Chen Y, Xiao J (2010). A 1-bp deletion in the gammaC-crystallin leads to dominant cataracts in mice. Mamm. Genome *21*, 361-369.

Zheng Q, Cai X, Tan MH, Schaffert S, Arnold CP, Gong X, Chen CZ, Huang S (2014). Precise gene deletion and replacement using the CRISPR/Cas9 system in human cells. Biotechniques *57*, 115-124.

Zhou J, Wang J, Shen B, Chen L, Su Y, Yang J, Zhang W, Tian X, Huang X (2014). Dual sgRNAs facilitate CRISPR/Cas9-mediated mouse genome targeting. FEBS J. *281*, 1717-1725.

Zhuang Y, Liu J, Wu H, Zhu Q, Yan Y, Meng H, Chen PR, Yi C (2022). Increasing the efficiency and precision of prime editing with guide RNA pairs. Nat. Chem. Biol. *18*, 29-37.

Zuo E, Cai YJ, Li K, Wei Y, Wang BA, Sun Y, Liu Z, Liu J, Hu X, Wei W, Huo X, Shi L, Tang C, Liang D, Wang Y, Nie YH, Zhang CC, Yao X, Wang X, Zhou C, Ying W, Wang Q, Chen RC, Shen Q, Xu GL, Li J, Sun Q, Xiong ZQ, Yang H (2017). One-step generation of complete gene knockout mice and monkeys by CRISPR/Cas9-mediated gene editing with multiple sgRNAs. Cell Res. *27*, 933-945.