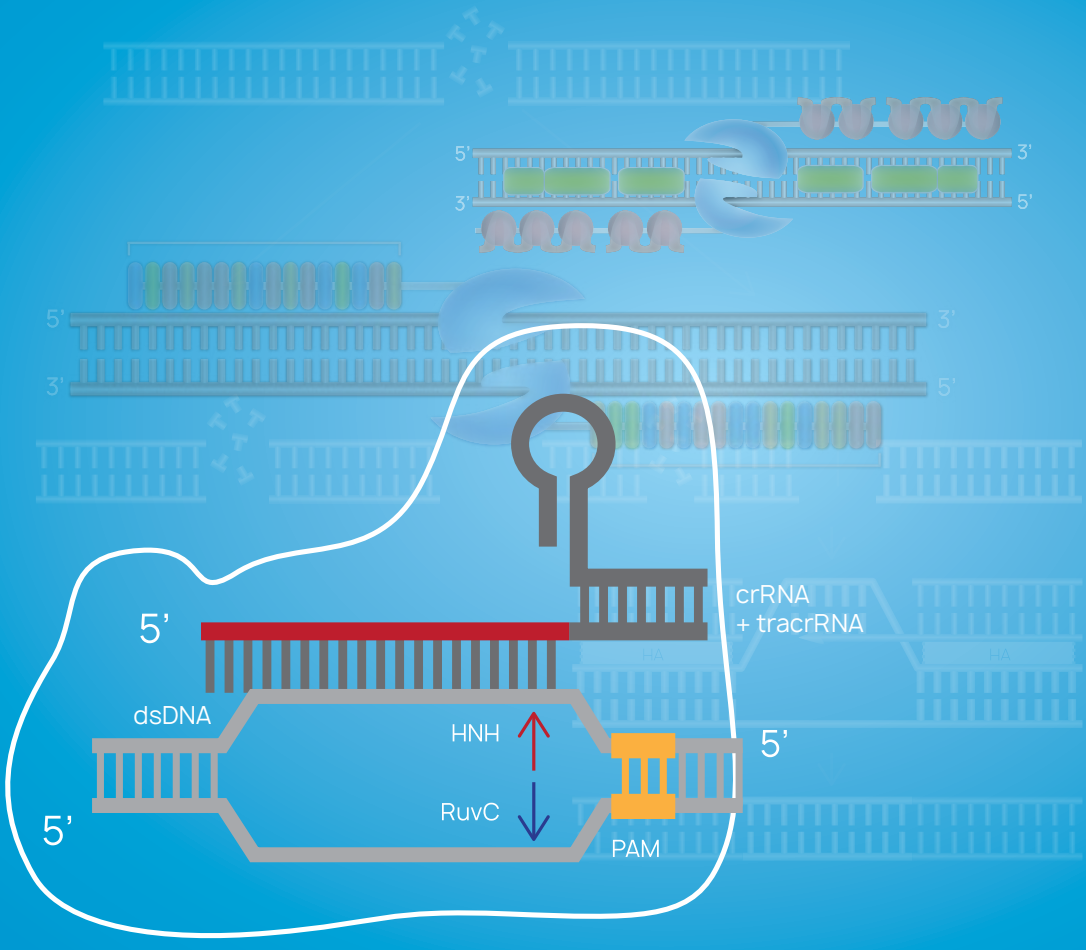
GENETIC TECHNOLOGIES



Central Research Institute for Epidemiology

Federal Service for Surveillance
on Consumer Rights Protection and Human Wellbeing
Central Research Institute for Epidemiology of Rospotrebnadzor

GENETIC TECHNOLOGIES

Edited by
Vasily G. Akimkin
Full Member of the Russian Academy of Sciences

Moscow
Central Research Institute for Epidemiology
2021

Translated from Russian

Генетические технологии / Ю.В. Михайлова, А.М. Нагорных, В.В. Петров, А.Е. Судьина, А.И. Тюменцев, М.А. Тюменцева, А.А. Шеленков; под общей редакцией академика РАН В.Г. Акимкина. — М.: ФБУН ЦНИИ Эпидемиологии Роспотребнадзора, 2020. — 200 с.
ISBN 978-5-6045286-0-0
DOI: https://doi.org/10.36233/978-5-6045286-0-0

Reviewers:
Victor V. Maleev, Full Member of the Russian Academy of Sciences, Dr. Sci. (Medicine), Professor, laureate of the State Prize and the Prize of the Government of the Russian Federation of science and technology, advisor to the director of the Central Research Institute for Epidemiology of Rospotrebnadzor for science
Areg A. Totolian, Full Member of the Russian Academy of Sciences, Dr. Sci. (Medicine), Professor, director of the Saint-Petersburg Pasteur Institute

Genetic Technologies / Yu.V. Mikhaylova, A.M. Nagornykh, V.V. Petrov, A.E. Sud'ina, A.I. Tyumentsev, M.A. Tyumentseva, A.A. Shelenkov; edited by V.G. Akimkin, RAS Full Member; [translated from Russian]. — Moscow: Central Research Institute for Epidemiology, 2021. — 176 p.

The book 'Genetic Technologies' covers a wide range of topics in the field of genetic technologies related to genome editing, and summarizes the latest scientific data on the use of programmable nucleases in various fields related to genome modification. The monograph 'Genetic Technologies' will be of interest and benefit to students of biological and medical universities, graduate students, young scientists, researchers and healthcare professionals and will help them to gain and align knowledge in the rapidly developing field of genetic technologies and genome editing. It is believed that the innovative technologies described in the book will serve as a scientific basis for the development of diagnostic and therapeutic agents to combat currently incurable diseases.

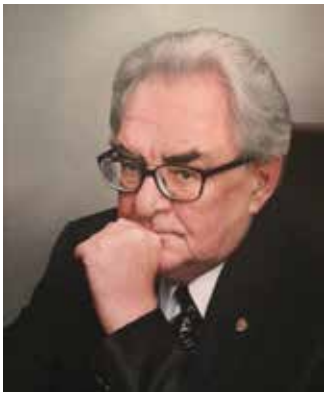
The book 'Genetic Technologies' was published with the support of the Ministry of Science and Higher Education of the Russian Federation under a grant for the creation and development of the 'World Class Genomic Research Center for Ensuring Biological Safety and Technological Independence under the Federal Scientific and Technical Program for the Development of Genetic Technologies' agreement No. 075-15-2019-1666.

DOI: https://doi.org/10.36233/978-5-6045286-3-1
ISBN 978-5-6045286-3-1

Abbreviations

CRISPR — clustered regularly interspaced short palindromic repeats
DMD — Duchenne muscular dystrophy
HBV — hepatitis B virus
iPSCs — induced pluripotent stem cells
MPS — mucopolysaccharidosis
NGF — nerve growth factor
NGS — next-generation sequencing
NHEJ — non-homologous end joining
NHPs — nonhuman primates
NLS — nuclear localization signal
PBS — primer binding site
RVD — repeat variable diresidue
TALE — transcription activator-like effectors
TALEN — transcription activator-like effector nucleases
ZFN — zinc-finger nucleases

TABLE OF CONTENTS



Preface

For thousands of years, people have sought to develop and improve beneficial traits of different organisms. Studies showed that changes in the genetic code of an organism would produce changes in the organism itself. In the late 1920s, before DNA was discovered, Lewis John Stadler exposed plants to X-rays to induce random mutations. Random mutagenesis and selective breeding paved the way for further scientific achievements. Yet, it was the discovery of the DNA molecular structure in 1953 by James Watson and Francis Crick that marked a milestone in the design and genetic modification of organisms. The subsequent discovery of DNA restriction enzymes and development of transformation methods were successfully used by Rudolf Jaenisch in 1974 for creating the first genetically modified mouse.

The first recombinant DNA molecule containing SV40, bacteriophage and *Escherichia coli* DNA fragments was created in 1972 by Paul Berg, Stanley Cohen and Herbert Boyer, together with their colleagues. Genetically modified bacteria were created in 1976 and 1978 to produce such important human proteins as somatostatin and insulin.

All these scientific breakthroughs and discoveries opened a door to genetic engineering or genetic modification. Genetic engineering involves a set of techniques used for *in vitro* construction of recombinant DNA molecules to be further inserted into an organism to produce new genetic combinations. Genetic engineering aims to create cells (first of all, bacterial cells) capable of mass-producing *human* proteins. Genetic engineering makes it possible to overcome interspecific barriers and to transfer heritable traits from one organism to another, thus being widely used in plant and animal breeding.

The vast arsenal of tools used in genetic engineering includes polymerases for DNA amplification, restriction enzymes, restriction endonucleases, and ligases for molecular cloning and creating recombinant DNA molecules as well as many other enzymes and

proteins to work with small and medium-size DNA molecules. Yet, this versatile set of tools was of little help in the handling of large and complex genomes of higher organisms. The problem is that restriction enzymes can recognize only specific, relatively short DNA sequences. While such limited recognition capability is sufficient for viral and bacterial DNA, specific short nucleotide sequences are quite rare in bacterial DNA, it turns into an obstacle when plant and animal genomes are involved. Such genomes contain numerous short nucleotide sequences, which are recognized by restriction enzymes, thus excluding the possibility of targeting a particular site. Top-priority tasks set by biotechnology and fundamental medicine required efficient tools and precision targeting of specific DNA sites within genomes of higher organisms, including humans.

The following years were rich in impressive genetic modifications, which were technically challenging, expensive and time-consuming. Looking for more efficient methods, scientists have developed new tools and opened a new avenue in genetic engineering — gene editing.

Gene editing or genome editing is a type of genetic engineering, in which DNA is inserted, deleted or replaced in the genome of a living organism by using programmable, engineered nucleases or *molecular scissors*. Targeted genome editing with engineered nucleases is a technique used to modify DNA efficiently and precisely through double-strand breaks induced by highly specific nucleases at target sites and subsequent repair at gene loci after the intended changes are made. Nuclease-induced breaks can be repaired through one of the two pathways: non-homologous end-joining or homology-directed repair.

Site-specific double-strand DNA breaks are typically induced by using such engineered nucleases as zinc-finger nucleases, transcription activator-like effector nucleases, and CRISPR/Cas system-associated proteins (where CRISPR means clustered regularly interspaced short palindromic repeats).

Targeted genome editing with engineered nucleases has quickly become a powerful tool among genomic modification techniques. It has been successfully used in functional genomics focusing on the identification of functions of genes and genetic elements regulating gene expression as well as on decoding of mechanisms coordinating the operation of genes in a cell. In addition, engineered nucleases are frequently used for the validation of target genes associated with human diseases. They are used to generate gene knockout in several cell lines. Engineered nucleases are important in achieving a complete knockout of genes that resist RNA interference, which is another commonly used method of functional genomics.

In addition to functional genomics, engineered nucleases have found successful application in cell screening, which is used for creating modified cell lines with embeddable promoters, tags and reporter elements integrated into genes or intergenic regions. Engineered nucleases are frequently used to create and optimize cell lines with specified functions, for example, cell lines overexpressing proteins or antibodies for biotechnical and pharmaceutical applications.

Since 2009, when the first knockout rat was created, engineered nucleases have been successfully used at the organism level mostly to create animal models of human diseases and to improve crop varieties and livestock breeds.

Finally, engineered nucleases are used for developing therapeutic drugs. Clinical studies evaluating zinc-finger, CRISPR/Cas and TALE-nuclease-based therapeutic candidates were conducted in 2014, 2016 and 2017, respectively.

The authors attempted to cover a wide range of topics in genetic engineering dealing with targeted genome editing, and hope that this monograph will find its reader and will be useful for students majoring in biology and medicine as well as for young scientists. This monograph can also help researchers, healthcare administrators and workers, university undergraduate and postgraduate students gain an insight into rapidly developing genetic engineering technologies and targeted genome editing.

Full Member of the Russian Academy of Sciences, laureate of the State Prize and the Prize of the Government of the Russian Federation in science and technology, laureate of personalized awards of the Academy of Medical Sciences of the USSR and the Russian Academy of Sciences, Dr. Sci. (Medicine), Professor Valentin I. Pokrovsky

CHAPTER 1
Artificial zinc-finger nucleases

Most of the approaches in the targeted modification of DNA sequences in the genome involve a double-strand break created at a specific DNA location and further repaired by using one of the two alternative mechanisms. Non-homologous end-joining (NHEJ), a quick ligation of DNA ends, may result in restoration of the initial DNA structure or in introducing mutations like insertions or deletions of varying lengths, which, in their turn, may adversely affect gene transcription [1]. This pathway can be used to knock out the required target gene. Homologous recombination occurs in the presence of homologous DNA of endogenous or exogenous origin. Thus, homologous recombination can be used for introducing targeted point mutations (for example, for genome correction) or for introducing the desired sequence through recombination of the target and the DNA matrix [2].

The first truly efficient tools in targeted genome editing were chimeric (hybrid) nucleases with specified action [3]. Such nucleases consist of two domains, one of which catalyzes DNA cleavage, while the other can bind selectively to specific nucleotide sequences in DNA.

Currently, genetic engineering offers four main types of hybrid nucleases, which can be programmed to induce a double-strand break [4–6]:

- meganucleases — highly specific homing endonucleases recognizing sequences of more than 14 base pairs of nucleotides. The term *meganucleases* most often implies homing endonucleases of the LAGLIDADG family;
- zinc-finger nucleases (ZFNs) [7];
- transcription activator-like effector nucleases (TALENs);
- clustered regularly interspaced short palindromic repeats (CRISPR).

This chapter focuses on the history and structure of nucleases with a zinc-finger DNA-binding domain and outlines their main characteristics and possible applications in targeted genome editing.

1.1. History of zinc-finger nucleases

The first reports about a new chimeric protein consisting of a FokI restriction endonuclease cleavage domain and a DNA-binding zinc-finger domain date back to 1996 [8]. These hybrid nucleases owe their birth to the discovery of a new type of DNA-binding domains in 1985 during studies of the transcription factor TFIIIA in *Xenopus oocytes* [9]. The domains were later named *zinc fingers*, as each domain (finger) contains a zinc ion [10]. In 1991, the analysis of the three-dimensional structure of the Zif268 transcription factor-DNA complex demonstrated a direct interaction between each zinc finger and specific DNA triplets [11]; later, a code for DNA sequence recognition by zinc-finger proteins was defined [12, 13].

In 1992, studies of the FokI restriction endonuclease showed that the DNA-binding and nuclease domains can act independently of each other [14]. The assumption suggesting that the action of the nuclease domain can be re-directed to another restriction site by replacing the DNA-binding protein domain was later confirmed [15] and provided another factor to push forward the construction of hybrid nucleases. The subsequent studies and work on hybrids of the FokI restriction nuclease domain and different DNA-binding domains, including zinc fingers, proved the possibility of targeted programming of nuclease specificity for such chimeric proteins [8, 16].

The first practical application of hybrid ZFNs was demonstrated in 2002 by using targeted genome editing in the fruit fly *Drosophila melanogaster* [17, 18]. Ever since, ZFNs have been used successfully for genome editing in more than a hundred genes of different plants [19–21], animals [22–25] and even human cell lines [26–28].

ZFNs have the best prospects in gene therapy. There are hybrid nucleases targeted at the correction of mutations causing different genetic disorders in humans: sickle cell disease [29, 30], hemophilia [31], various neurodegenerative diseases [32–34], and muscular dystrophies [35].

Another therapeutic application of hybrid ZFNs is gene therapy for HIV infection. The main approach involves disruption/removal of receptor genes, through which the virus enters T cells, thus preventing the virus from infecting lymphocytes and facilitating the recovery of the T-cell population [7, 36, 37]. The results were so successful that the CCR5-specific nuclease or SB-728 gene therapy of HIV infection (Sangamo Therapeutics) is currently tested in several clinical trials[1].

1.2. Structure and functions of zinc-finger nucleases

Zinc-finger nucleases (ZFNs) are fusion or hybrid proteins consisting of a non-specific FokI restriction endonuclease cleavage domain and a zinc-finger DNA-binding domain. The DNA-binding domain has a modular structure and typically contains 3–6 ZF motifs.

[1]URL: https://aidsinfo.nih.gov/clinical-trials/search?q=ZFN&c=clinicaltrials&curID=161581

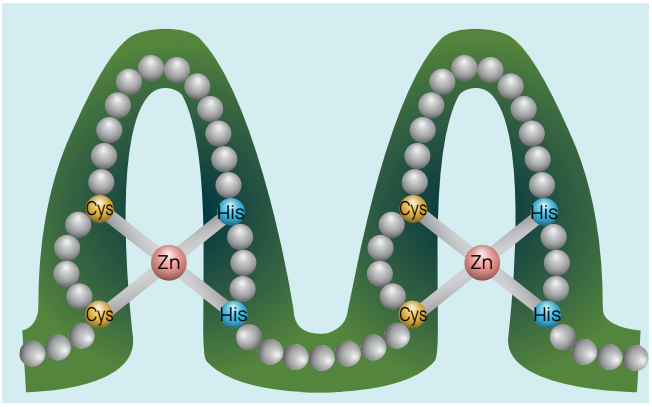


Fig. 1.1. Structure of C2H2 zinc-finger proteins

ZF DNA-binding domain

C2H2 ZFs are the most common type of zinc fingers, representing a module with approximately 30 amino acids, which coordinate 1 zinc ion with 2 cysteines and 2 histidines (Cys2His2). The C2H2 domain is found in many transcription factors. Each C2H2 domain folds into a compact ββα structure consisting of an antiparallel β-sheet followed by an α-helix, which is stabilized by a zinc ion **(Fig. 1.1)**. In the repeat, 25 of the 30 amino acids fold around a zinc ion to form a small independent structural domain, *a finger*, while the other 5 amino acids (TGEK(R)P) provide short linkers between consecutive fingers [38]. The zinc ion is tetrahedrally coordinated between 2 cysteine and 2 histidine residues to stabilize the fingers.

Structural analyses of ZF proteins and resulting findings helped identify key points in the ZF and DNA interaction. As shown in the crystal structure of Aart ZF protein, ZF α-helices bind in the major groove of the DNA **(Fig. 1.2)** [39]. The Aart ZF protein is

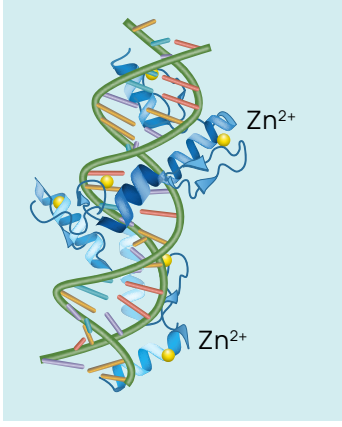


Fig. 1.2. Structure of the Aart ZF protein and DNA complex (PDB ID 2l13)

an artificial protein assembled from 6 ZF motifs interconnected with linkers consisting of 5 amino acid residues. The Aart ZF protein recognizes 18 bp sequences of DNA.

Each *finger* selectively recognizes three nucleotides in the DNA. In most cases, the interaction is provided by three amino acid residues located at positions -1, 3 and 6 within the α-helix and come into contact with the recognized 3′-end, middle and 5′-end triplet nucleotides, respectively **(Fig. 1.3)**. Some ZFs can also have a cross-strand interaction between amino acid residues at position 2 on the α-helix and nucleotides of the opposite strand of the double helix. Besides, amino acid residues at positions -2, 1 and 5 within the α-helix make direct or water-mediated bonds with the phosphate backbone of DNA [40].

It should be noted that the module containing several ZFs encodes a continuous recognition DNA site consisting of the respective triplets [41]. Thus, a combination of specific ZF motifs makes it possible to program recognition of virtually any sequence in a DNA. Several approaches have been used to identify ZFs for each of 64 possible DNA triplets and to create a library of specific ZF motifs: a standard search among the known ZF proteins [42, 43] as well as a selection of synthetic Zif268 variants by using the phage display technique [44, 45]. Multi ZFs can be constructed from the known ZFs assembled in any order to recognize any desired DNA sequence by using the modular assembly method [46]. Many of the ZF domains constructed by modular assembly are characterized by a higher specificity as compared to natural ZF domains [47]. Being one of the most preferred methods of constructing ZF proteins, modular assembly is widely used for creating hybrid nucleases, transposases, recombinases, integrases, etc [48–50].

In the meantime, the currently known ZF modules can recognize only all GNN triplets, most ANN and CNN triplets, and a few TNN triplets (where N can be any of the four nucleotides) [51–53]. The above circumstance, as well as the cross-strand interaction of ZFs with the second DNA strand and the possible impact of neighboring ZFs on the specificity of recognition of target triplets, impede the creation of a complete library of ZF modules [54, 55].

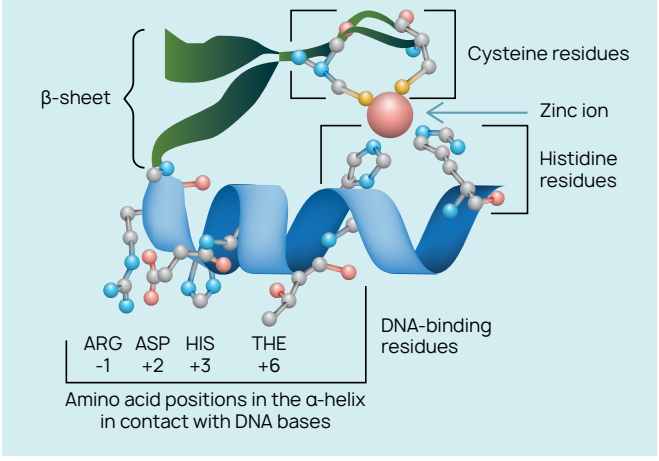


Fig. 1.3. Design of the C2H2 zinc-finger domain

Prospects of creating programmable ZF proteins spurred the launching of two business companies in the late 1990s. Sangamo BioSciences was founded by Edward Lanphier in 1995; initially, it focused on designing ZF-based synthetic transcription factors. The other company, Gendaq Ltd, was founded by Aaron Klug and Yen Choo in 1999 to specialize in ZF libraries of a broad specificity spectrum. In 2001, Sangamo BioSciences acquired Gendaq Ltd and its ZF module database. In the next years, Sangamo BioSciences made significant progress in creating hybrid ZF transcription factors and ZFNs for gene therapy, including treatment of HIV infection.

FokI nuclease domain

In chimeric ZFNs, the nuclease domain is frequently represented by the catalytic domain of FokI restriction enzyme, which belongs to type IIS restriction endonucleases. Enzymes of this type recognize an asymmetric nucleotide sequence and cleave at least one strand of the DNA substrate at some distance from the recognition site. The cleavage of the second strand can occur both inside and outside the recognizable sequence [56]. Most of the enzymes of this type cleave both DNA strands outside the recognition site, thus leaving them intact.

FokI is the most extensively studied representative of this group [56, 57]. The structure of FokI restriction endonuclease has two domains: the N-terminal domain is DNA-binding, while the C-terminal domain is responsible for catalysis. It was shown that divalent metal ions are dimerization cofactors and that the active complex requires coordination of two recognition sites [58]. FokI α-helices forming a dimerization module are located in the C-terminal catalytic domain of the enzyme **(Fig. 1.4, *a*)** [59]. During the first stage, the FokI monomer binds to the recognition site in the DNA. Once the first monomer is bound to the recognition site, the second FokI monomer binds to the site due to the interaction of two DNA-cleavage domains. The DNA-binding domain of the second monomer can be located in the solution or can be bound to the other recognition site in the DNA [60]. The coordinated action of two protein molecules induces a double-strand break in the DNA **(Fig. 1.4, *b*)** [61]. The monomer that was the first to bind to the recognition site induces a single-strand break in the bottom strand of the DNA, while the second monomer targets the top strand of the DNA [62]. Thus, in the case of FokI, dimerization is required for creating the active center of the enzyme [57, 61].

ZFN design

Highly specific zinc-finger chimeric nucleases are generated by fusing a zinc-finger DNA-binding domain to a FokI DNA-cleavage domain. The FokI nuclease domain does not contribute to the substrate specificity of ZF nuclease, but it must dimerize to cleave DNA efficiently [61]. Such dimerization can be achieved by using a pair of chimeric proteins consisting of FokI and ZF domains recognizing the non-palindromic DNA sequence. The FokI nuclease domain is linked to the C-terminus of the ZF domain. The ZF strand consists of single or double-finger motifs linked through spacers. The number of ZF motifs in each domain can range from three to six. ZF recognition sites must be located in opposite DNA strands

in an inverted orientation to each other and separated by 5–7 bps. It is important for the proper orientation of the active centers of FokI domains and the maximum cleavage efficiency **(Fig. 1.5)** [63].

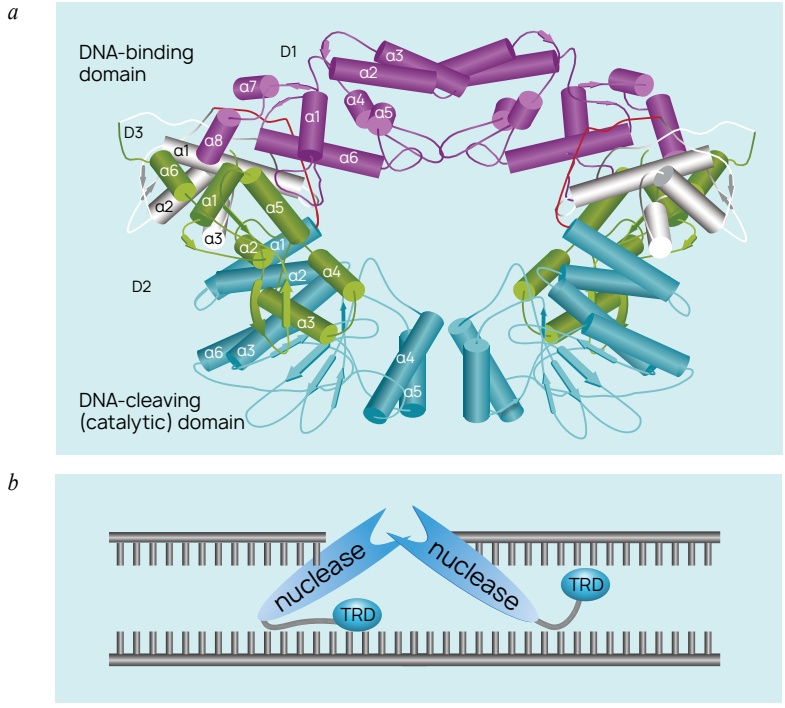


Fig. 1.4. Structure of the FokI restriction endonuclease dimer (*a*) and the schematic representation of interaction between the FokI restriction endonuclease and the DNA (*b*).
a: the catalytic domain is indicated in blue color; the recognition domain is indicated in purple.
b: TRD (target recognition domain) — the domain responsible for substrate specificity

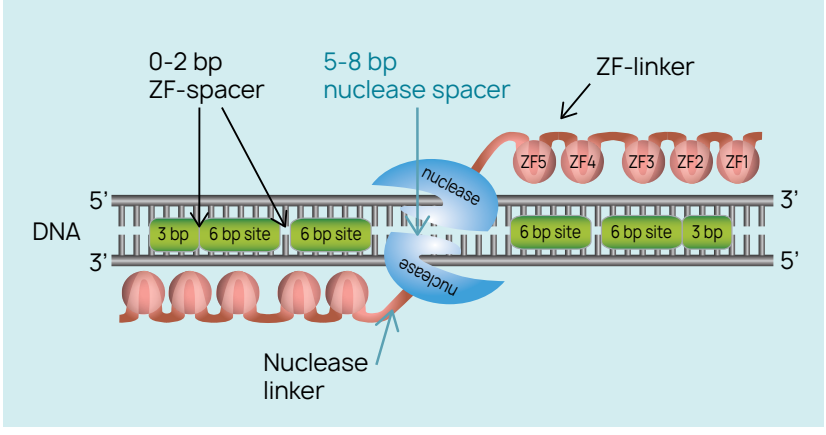


Fig. 1.5. Schematic representation of the standard ZFN model

Considering some degeneracy of the recognition code, the application of heterodimeric ZFNs can cause an unwanted reaction that can be potentially toxic for cells — cleavage of off-target genome sites by homodimers consisting of two *left* or *right* nucleases. Furthermore, the increased and unrepairable off-target cleavage leads to unwanted mutations or cell death [17, 22, 64]. The above downsides can be avoided by replacing some amino acid residues in a nuclease domain and placing an emphasis on heterodimerization rather than homodimerization [65–68]. This would not only result in considerably reduced ZFN cytotoxicity being a major issue with ZFN therapeutic applications but also demonstrate absent nuclease activity at locations of potential palindromes for homopolymeric ZFNs [67, 69–72].

1.3. Application of zinc-finger nucleases

DNA double-strand break repair

The DNA-binding domain enables highly specific binding of ZFNs to the target DNA sequence up to 18 bp in length, thus making them an efficient tool for targeted genome editing, which can be used to induce a DNA double-strand break at a specified locus within a large genome. The DNA damage triggers the activation of repair systems. **Fig. 1.6** shows existing repair mechanisms of double-strand breaks.

NHEJ rejoins broken DNA ends by direct ligation and, in contrast to homologous recombination, does not require any donor DNA. The term 'non-homologous end joining' was coined in 1996 by J.K. Moore and J.E. Haber [73]. NHEJ offers significantly lower precision compared to homologous recombination. Such imprecise repair leads to loss of nucleotides and small insertions or mutations. The NHEJ repair can proceed in two ways. In the first case, after a break has been induced, the direct

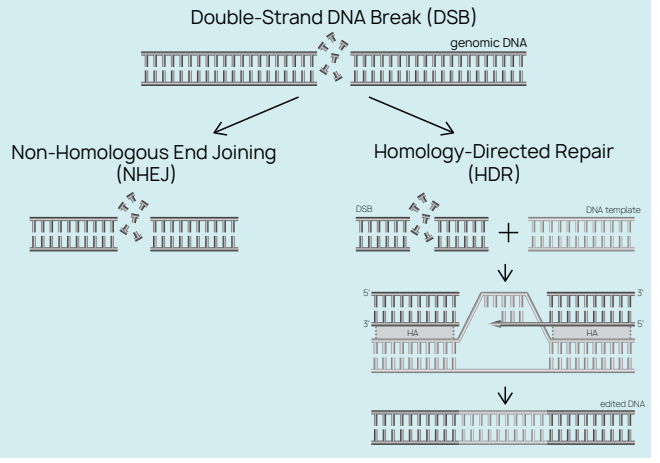


Fig. 1.6. Repair mechanisms of a DNA double-strand break

restoration (ligation) of the sequence at the target site results in a gene knockout caused by the insertion or deletion of nucleotides (indels). In the other case, two concurrently used pairs of ZFNs may cause long deletions [28].

NHEJ is the main pathway for the repair of double-strand breaks in DNA [74]; it is widely used in ZFN applications, including practical genetic engineering and therapy [37]. It is typically used in the disruption of coding sequences (a gene knockout) or reading-frame restoration **(Fig. 1.7)**. In the meantime, NHEJ is an error-prone and uncontrollable process, and cannot be used for targeted genome modification.

On the other hand, homologous recombination offers precise editing of the target sequence in the presence of a homologous chromosome or exogenous DNA **(Fig. 1.6)** [75, 76]. This approach is recommended for target-specific DNA modification. As shown in Fig. 1.7, depending on the design of the donor DNA, when ZFNs are introduced along with the donor DNA, the genetic mutation can be removed through correction of single nucleotides, the introduction of long fragments (transgenes)

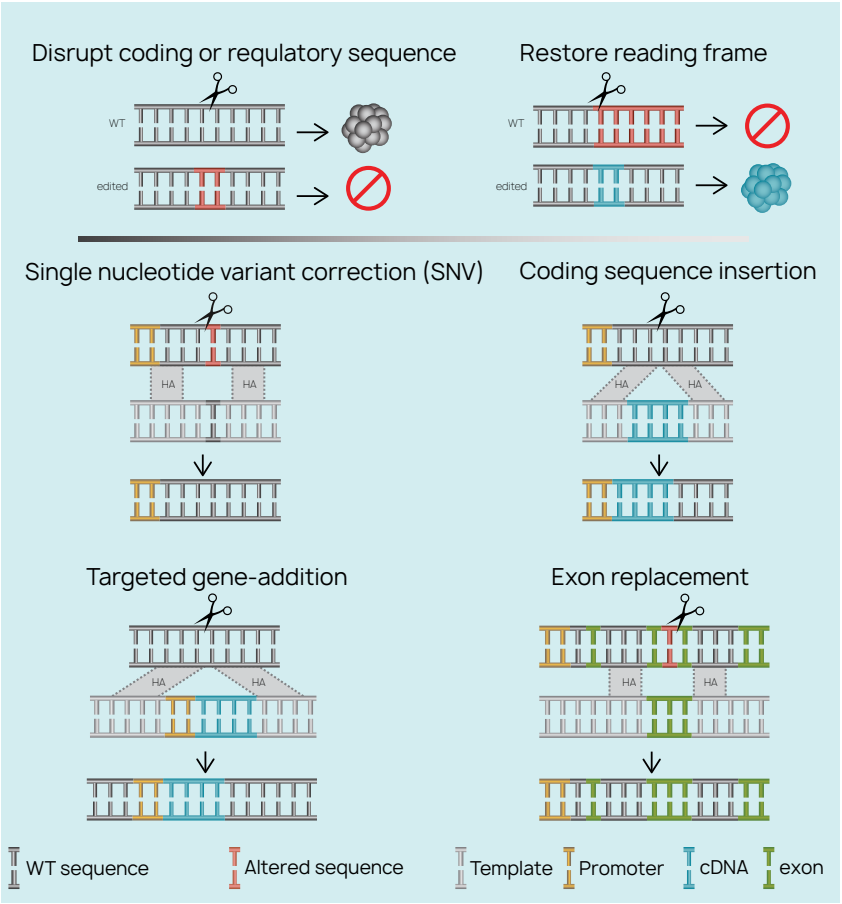


Fig. 1.7. Types of genome modification and their therapeutic applications

into DNA in the required orientation, addition of genes (a promoter + cDNA) to the required locus and replacement of coding sequences (exons). Low modification of the target DNA, a major drawback caused by using ZFNs for induction of homologous recombination, limits the application of this approach in gene therapy and requires enrichment of modified cells [77].

Development and application of ZFNs in therapy

Since its first steps in the early 2000s, ZFN technology has matured greatly. It is used for the modification of genomes of diverse plant and animal organisms. Despite their universal design, ZFNs have several disadvantages strongly limiting their application, especially in clinical practice. Possible solutions include:

- increased specificity and minimization of side effects through new approaches to the design;
- improved efficacy and safety of application, and reduced immunogenicity through improved methods of delivery.

The construction of ZFNs is still a labor-intensive, time-consuming, and costly process. The major challenge is posed by a DNA-binding domain, which is responsible for the specificity of DNA sequence recognition. In addition to the limited set of ZF domains, other downsides are the lack of a universal algorithm of ZF modular assembly design and existing position-dependent effects of neighboring ZF modules [78]. Therefore, the development of approaches and solutions for designing ZF modules with specified characteristics and high substrate specificity generates a lot of interest. These approaches include design of the most suitable modular assemblies consisting of 3–6 ZFs and subsequent selection employing different techniques, such as bacteriophages or bacterial two-hybrid selection. Oligomerized pool engineering (OPEN) [54] and context-dependent assembly (CoDA) [79] are the most common methods used in ZFN construction [79].

The required high ZFN specificity resulted in the prevalence of nucleases containing 4–6 ZFs. As mentioned previously, binding of ZF nucleases is sequence-specific. However, such long binding sites are characterized by a certain similarity between target and off-target sequences, thus posing a risk of off-target DNA cleavage or the so-called off-target effect [80, 81]. It was found that off-target DNA cleavage could take place at 66% homology [82]. Such off-target ZFN activity can cause DNA impairment resulting in cytotoxicity, apoptosis and major chromosomal rearrangements in cells [83]; the higher the activity is, the longer the recognition site will be. Therefore, when creating artificial ZFNs, researchers should find a reasonable balance between achieving high specificity through increased numbers of ZF modules and diminishing off-target-effects through the reduction in modules.

Another factor contributing to the increased cleavage frequency for off-target DNA sequence is the ability of ZFNs to act as homodimers [84]. To reduce this effect, the nuclease domain was modified to give priority to heterodimerization or to turn it into the only way of creating an active ZFN dimer [67].

Efficacy and safety of genome editing and, consequently, of therapy is limited by the delivery of candidate molecules directly to target cells [85, 86]. Approaches

to delivery of the therapeutic agent for genome editing can be divided into *in vivo* and *ex vivo* strategies **(Fig. 1.8)**.

The employed approach to delivery will depend on the type of a tool for genome editing. Hybrid nucleases are delivered to a cell as genetically engineered constructs to accumulate the respective proteins within the cell **(Fig. 1.9)**. The delivery system must ensure highly efficient penetration of these gene constructs into the cell, resistance to degradation in the cell during transportation to the nucleus, and maintenance of the required level of expression. For example, when the *in vivo* strategy is employed, genome editing tools will be affected by the host immune system. The potential immune response will depend on the type of the delivery vehicle. Application of

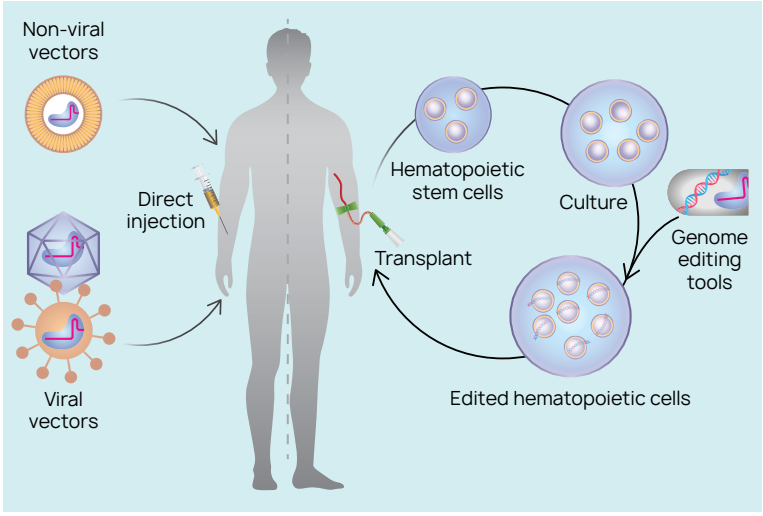


Fig. 1.8. Delivery strategies for therapeutic agents intended for genome editing.
iPSCs — induced pluripotent stem cells

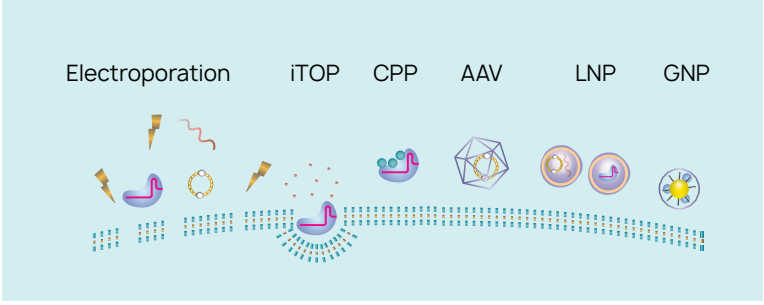


Fig. 1.9. Physical and chemical methods of delivery of genetically engineered constructs to cells of an organism:
electroporation; iTOP — induced transduction; CPP — cell-penetrating peptides; AAV — adeno-associated virus, including vectors; LNP — lipid nanoparticles; GNP — gold nanoparticles

viral vectors can lead to long-term expression of hybrid nucleases, which, in its turn, can cause extensive damage to a human genome and prolonged immune response.

Numerous viral and non-viral systems for delivery of genetically engineered constructs to cells of an organism have been developed lately [87, 88]. The most widely used viral systems are systems based on retroviruses, lentiviruses, adenoviruses, adeno-associated viruses, and the herpes simplex virus. For example, different serotypes of adeno-associated viruses enhance the efficiency of delivery for certain types of cells, thus making it tissue-specific [89]. The *in vivo* delivery of hybrid nucleases with adeno-associated viruses has been successfully used on different animal models involving metabolic disorders [90, 91], infection caused by the human immunodeficiency virus [92], muscular dystrophies [93], retinal diseases [94, 95], neurodegenerative diseases [96], *etc*.

As viral vectors contribute to the efficient delivery and the longer-lasting expression of genome editing tools, they are increasingly promising for clinical application. However, immune responses induced by viral delivery systems can be a crucial factor in limiting the therapeutic potential of the delivered construct for genome editing [97, 98]. It turns into a major challenge when genome editing implies repeated (or multiple and long-term) administration of a therapeutic gene product. Some of these limitations can be overcome with combined immunosupressive treatment [99].

Non-viral systems involve direct administration of DNA/RNA to cells and tissues by using electroporation, liposomes, cationic polymers, *etc*. [100]. Recently developed lipid-based nanoparticles have been approved for therapeutic application [101]; gold nanoparticles were successfully used in rodent models to treat Duchenne muscular dystrophy [102] and fragile X syndrome [103].

Similar delivery systems are also used *ex vivo* for precise genome editing. Depending on the type of cells used in *ex vivo* delivery, tools for genome editing can be delivered both through viral vectors and by using electroporation, microinjections, cell-penetrating peptides or nanoparticles. Pluripotent stem cells are also popular in *ex vivo* genome editing applications [104]. Induced pluripotent stem cells (iPSCs) have attracted considerable interest as promising model systems, as they can be differentiated into any type of cells relating to the studied disease, for example, into skeletal muscle cells [105–107], hepatocytes [108, 109], cardiac muscle cells [110, 111], and many others.

Each of these delivery formats has its advantages and drawbacks; therefore, in most cases, preference is given to a combination of viral and non-viral systems [112]. Besides, it has been demonstrated that ZFNs possess the innate ability to cross cell membranes and induce a targeted gene knockout in human cells [113]. In direct ZFN delivery, the off-target DNA cleavage rates were significantly lower than the rates observed for ZFN expression in the cell. Lower rates of off-target genome cleavage can be caused by the shorter time, during which the nuclease stays in the cell. When this approach is applied, high rates of the gene knockout can be reached only after repeated ZFN treatment of cells, thus substantially limiting the application of this technique for *ex vivo* genome editing. Nevertheless, incorporation of tandem NLS repeats (where NLS mean nuclear localization signals) into the ZFN backbone can increase ZFN cell-penetrating activity by up to 13 times [114]. Furthermore, even one-time treatment enhanced the efficacy of the gene knockout in many types of human cells, including CD4+ T cells and iPSCs.

The rates of genome modification by multi-NLS ZFNs delivered directly into cells exceeded the rates that were achieved with viral delivery vectors or plasmid DNA.

ZFNs in clinical trials

In recent years, researchers have been looking for possible applications of hybrid nucleases (ZFNs, TALENs and CRISPR/Cas9) in genome editing. Their therapeutic applications include gene therapy of HIV-infection, cancer, and genetic disorders. **Table 1.1** presents summarized information about ZFN therapeutic products going through different phases of clinical trials (Source: clinicaltrials.gov).

The first in-human application of targeted genome editing with ZFNs for *ex vivo* cell therapy involved the disruption of the CCR5 gene in CD4+ T cells of HIV patients (NCT00842634) [37]. CCR5 is a transmembrane β-chemokine receptor expressed on the surface of activated T-cells, the main co-receptor for HIV-1 entry [115]. The discovery of the 32-nucleotide deletion in the exon of CCR5 gene (CCR5-Δ32) and resulting mutant protein not expressed on the surface of T cells in individuals with natural resistance to HIV-1 infection helped find the ideal solution in gene therapy of HIV infection — a knockout of the CCR5 receptor similar to the Δ32 genotype [116, 117]. The preclinical studies identified a ZFN pair (later named SB-728) capable of producing a double-strand break within a transmembrane domain, upstream the naturally occurring CCR5-Δ32 mutation typical of human CD4+ T-cells, and generating deletions, which prevent the expression of the *CCR5* gene [36]. Later on, in 2009, Sangamo Therapeutics initiated the first clinical trial (NCT00842634) for evaluation of therapeutic safety of autologous CD4+ T cells modified *ex vivo* by using ZFN SB-728. The adenoviral vector system was used for delivery. The results demonstrated that edited T cells are safe for patients; in addition, partially acquired resistance to HIV was reported; the detected HIV DNA levels decreased in most of the patients, thus proving the efficacy of this gene therapy [37]. At present, the knockout of CCR5-receptor in cells of HIV patients is the most popular approach in therapeutic genome editing.

At the moment, a total of ten HIV-targeted ZFN-modified therapeutic products are being tested and evaluated through clinical trials (**Table 1.1**). The versions of the initial approach included:

- escalation of the edited cell dose in infusion (NCT01044654);
- lymphodepletion (with cyclophosphamide) before infusion of edited T cells to reduce the number of unedited T cells and improve infused T cell engraftment (NCT01543152);
- repeated infusion of edited cells (NCT02225665);
- using mRNA electroporation for ZFN delivery into T cells (NCT02388594);
- using modified ZFN autologous CD34+ CCR5-negative hematopoietic stem and progenitor cells (NCT02500849).

Preclinical studies found that ZFNs could be used as a knockout of CCR5 in CD34+ hematopoietic stem and progenitor cells, thus generating a CCR5-negative progeny [118]. In addition, mRNA electroporation used for ZFN delivery has lower cytotoxicity compared to adenoviral vectors and can be more easily scaled up to reach the levels required for treatment [119].

Table 1.1. List of ZFN-based genome editing therapeutic candidates at different phases of clinical trials

Condition or disease	Study title	Study start and completion dates	Drug description	Developer	*Clinicaltrials.gov* Identifier
HIV infection	A Phase I Study of Autologous T-Cells Genetically Modified at the CCR5 Gene by Zinc Finger Nucleases SB-728 in HIV-Infected Patients	January 2009 – January 2013	SB-728-T — autologous CD4+ T cells modified *ex vivo* by ZFN SB-728; AdV delivery	University of Pennsylvania	NCT00842634
	A Phase 1 Dose Escalation, Single Dose Study of Autologous T-Cells Genetically Modified at the CCR5 Gene by Zinc Finger Nucleases SB-278 in HIV-Infected Patients Who Have Exhibited Suboptimal CD4+ T-Cell Gains During Long-Term Antiretroviral Therapy	December 2009 – December 2014	SB-728-T — autologous CD4+ T cells modified *ex vivo* by ZFN SB-728; AdV delivery	Sangamo Therapeutics	NCT01044654
	A Phase 1/2, Open Label, Single Infusion Study of Autologous T-Cells Genetically Modified at the CCR5 Gene by Zinc Finger Nucleases (SB-728-T) in HIV Infected Subjects	November 2010 – May 2015	SB-728-T — autologous CD4+ T cells modified *ex vivo* by ZFN SB-728; AdV delivery	Sangamo Therapeutics	NCT01252641
	Long-Term Follow-up of HIV-Infected Subjects Treated With Autologous T-Cells Genetically Modified at the CCR5 Gene by Zinc Finger Nucleases (SB-728-T or SB-728mR-T)	March 2011 – December 2031	Patients being administered SB-728-T or SB-728mR-T therapy	Sangamo Therapeutics	NCT04201782
	A Phase I, Open-Label Study to Assess the Effect of Escalating Doses of Cyclophosphamide on the Engraftment of SB-728-T in Aviremic HIV-Infected Subjects on HAART	December 2011 – July 2017	SB-728-T — autologous CD4+ T cells modified *ex vivo* by ZFN SB-728; AdV delivery	Sangamo Therapeutics	NCT01543152

Table 1.1 to be continued on page 23.

Condition or disease	Study title	Study start and completion dates	Drug description	Developer	*Clinicaltrials.gov* Identifier
	A Phase I Study of T-Cells Genetically Modified at the CCR5 Gene by Zinc Finger Nucleases SB-728mR in HIV-Infected Patients, With or Without the CCR5 Delta-32 Mutation, Pre-treated With Cyclophosphamide	April 2015 – March 2019	SB-728mR-T — autologous CD4+ T cells modified *ex vivo* by ZFN SB-728; electroporated SB-728 mRNA	University of Pennsylvania	NCT02388594
	A Pilot Study to Evaluate the Feasibility, Safety and Engraftment of Zinc Finger Nucleases (ZFN) CCR5 Modified CD34+ Hematopoietic Stem/ Progenitor Cells (SB-728mR-HSPC) in HIV-1 (R5) Infected Patients	July 2015 – April 2022	SB-728mR-HSPC — autologous CD34+ hematopoietic stem and progenitor cells modified *ex vivo* by ZFN SB-728; electroporated SB-728 mRNA	City of Hope Medical Center	NCT02500849
	T-Cell Reinfusion After Interfering With Lymphocyte Binding Location of AIDS Virus Through Zinc Finger Nucleases Elimination of CCR5 Receptors: The TRAILBLAZER Study	June 2019 – February 2024	SB-728-T — autologous CD4+ T cells modified *ex vivo* by ZFN SB-728; AdV delivery	Case Western Reserve University	NCT03666871
	A Pilot Study of T Cells Genetically Modified by Zinc Finger Nucleases SB-728mR and CD4 Chimeric Antigen Receptor in HIV-infected Subjects	July 2019 – December 2025	CD4 CAR+CCR5 ZFN T cells — SB-728mR-T cells expressing a CD4 receptor	University of Pennsylvania	NCT03617198
Transfusion-dependent β-thalassemia	A Phase 1/2, Open-label, Single-arm Study to Assess the Safety, Tolerability, and Efficacy of ST-400 Autologous Hematopoietic Stem Cell Transplant for Treatment of Transfusion-Dependent Beta-thalassemia (TDT)	March 2018 – March 2023	ST-400 — autologous CD34+ hematopoietic stem and progenitor cells modified *ex vivo*, targeting the *BCL11A* gene; electroporated ST-400 mRNA	Sangamo Therapeutics, Sanofi	NCT03432364

Table 1.1 to be continued on page 24.

Condition or disease	Study title	Study start and completion dates	Drug description	Developer	*Clinicaltrials.gov* Identifier
Sickle cell disease	A Phase 1/2, Open-Label, Multicenter, Single-Arm Study to Assess the Safety, Tolerability, and Efficacy of BIVV003 for Autologous Hematopoietic Stem Cell Transplantation in Patients With Severe Sickle Cell Disease	June 2019 – April 2023	BIVV003 — autologous CD34+ hematopoietic stem and progenitor cells modified *ex vivo*, targeting at the *BCL11A* gene; electroporated BIVV003 mRNA	Bioverativ, Sanofi	NCT03653247
Hemophilia B	A Phase I, Open-Label, Ascending Dose Study to Assess the Safety and Tolerability of AAV2/6 Factor IX Gene Therapy Via Zinc Finger Nucleases (ZFN) Mediated Targeted Integration of SB-FIX in Adult Subjects With Severe Hemophilia B	November 2016 – January 2021	SB-FIX — ZFN1, ZFN2 and donor cDNA	Sangamo Therapeutics	NCT02695160
MPS I	A Phase I / 2, Multicenter, Open-label, Single-dose, Dose-ranging Study to Assess the Safety and Tolerability of SB-318, a rAAV2/6-based Gene Transfer in Subjects With Mucopolysaccharidosis I (MPS I)	May 2017 – January 2022	SB-318 — ZFN1, ZFN2 and donor DNA (hIDUA)	Sangamo Therapeutics	NCT02702115
MPS II	A Phase I / 2, Multicenter, Open-label, Single-dose, Dose-ranging Study to Assess the Safety and Tolerability of SB-913, a rAAV2/6-based Gene Transfer in Subjects With Mucopolysaccharidosis II (MPS II)	May 2017 – February 2022	SB-913 — ZFN1, ZFN2 and donor DNA (hIDS)	Sangamo Therapeutics	NCT03041324

For several years, the using of the ZFN SB-728 product for the treatment of HIV infection (in various modifications) was the only precedent of ZFN application for genome editing at the stage of clinical trials. Nevertheless, the rapid development of TALEN technologies and especially CRISPR-associated nucleases have opened a door for prospective therapeutic ZFN agents for the treatment of different hemoglobinopathies caused by mutations in the β-globin gene, for example, in transfusion-dependent β-thalassemia or sickle cell disease. Sangamo Therapeutics, jointly with Bioverativ Therapeutics Inc, designed ZFNs targeted at the *BCL11A* gene, which represses the production of fetal hemoglobin in adults [120, 121]. Reactivation of fetal hemoglobin expression in adults is an efficient method of compensating for the *wrong* hemoglobin or a low β-hemoglobin level in sickle cell disease or transfusion-dependent β-thalassemia. Thus, ZFNs were aimed at genome editing for suppressing the *BCL11A* enhancer expression in autologous CD34+ hematopoietic stem and progenitor cells [122, 123]. When patients have such transformed cells infused, their levels of expression of endogenous hemoglobin increase, thus making blood transfusion in patients with transfusion-dependent β-thalassemia and sickle cell disease less challenging. At present, ZFN modified products, such as the ST-400 for treatment of transfusion-dependent β-thalassemia (NCT03432364) and the BIV003 for sickle cell disease (NCT03653247) are going through Phase I of the clinical trials.

Different levels of readiness for clinical application are also demonstrated by therapeutic products for the treatment of monogenic disorders associated with hemophilia B and Type I and II mucopolysaccharidosis (MPS). Hemophilia B, also known as Christmas disease, is a blood disorder caused by the deficiency of clotting factor IX. MPS I and II are metabolic connective tissue diseases caused by deficiencies of lysosomal enzymes required for the breakdown of glycosaminoglycans. The effect of products used in gene therapy of hemophilia B and MPS I and II is based on the injection of the adeno-associated virus 8 (AAV8) expressing ZFNs into the albumin locus in the liver cells, which secrete the deficient enzyme into the blood flow, for targeted integration of the normal copy of the gene into this locus [124]. Using this method, the authors were able to reach the physiological level of expression for some proteins, including α-galactosidase A (Fabry disease), clotting factor IX (hemophilia) and α-L-iduronidase, or IDUA (MPS II). The further testing of the targeted insertion of genes of iduronate-2-sulfatase (IDS) and IDUA enzymes into the albumin locus by using MPS I and II animal models demonstrated steadily increasing IDS and IDUA levels in blood [125]. At present, two ZFN products for the treatment of MPS I (NCT02702115) and II (NCT03041324) are being tested through clinical trials. It should be noted that both therapeutic products are only intended for patients with mild disease characterized by little to no damage to the central nervous system, as the enzymes produced by the liver cannot cross the blood-brain barrier. With the successful outcome, the ZFN SB-FIX product, which is going through Phase I clinical trial and is intended for patients with severe hemophilia B (NCT02695160), will help maintain the constant level of factor IX expression in the liver during the entire life of the patient.

Therefore, the expanding range of ZFN applications (including HIV therapy) and positive outcomes of Phase I/II clinical trials of therapeutic products for ZFN-media-

ted genome editing demonstrate that ZFNs are a prominent tool in clinical practice. Nevertheless, the transition to Phase III trials will require time and money for collecting, editing and expanding sufficient numbers of transformed cells for each patient.

References

1. Lieber M.R. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu. Rev. Biochem.* 2010; 79: 181-211.
2. Moynahan M.E., Jasin M. Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. *Nat. Rev. Mol. Cell Biol.* 2010; 11(3): 196-207.
3. Gaj T., Gersbach C.A., Barbas C.F. ZNF, TALEN and CRISPR/CAS-based methods for genome engineering. *Trends Biotechnol.* 2013; 31(7): 397-405.
4. Waryah C.B., Moses C., Arooj M., et al. Zinc fingers, TALEs, and CRISPR systems: A comparison of tools for epigenome editing. In: Jeltsch A., Rots M., eds. *Epigenome Editing. Methods in Molecular Biology. Volume 1767*. New York: Humana Press; 2018.
5. Pâques F., Duchateau P. Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy. *Curr. Gene Ther.* 2007; 7(1): 49-66.
6. Stoddard B. Homing endonucleases: from microbial genetic invaders to reagents for targeted DNA modification. *Structure.* 2011; 19(1): 7-15.
7. Urnov F.D., Rebar E.J., Holmes M.C., et al. Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.* 2010; 11(9): 636-46.
8. Kim Y.G., Cha J., Chandrasegaran S. Hybrid restriction enzymes: zinc finger fusions to FokI cleavage domain. *Proc. Natl. Acad. Sci. USA.* 1996; 93(3): 1156-60.
9. Miller J., McLachlan A.D., Klug A. Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. *EMBO J.* 1985; 4(6): 1609.
10. Klug A., Rhodes D. Zinc fingers: a novel protein fold for nucleic acid recognition. *Cold Spring Harb. Symp. Quant. Biol.* 1987; 52: 473-82.
11. Pavletich N.P., Pabo C.O. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A resolution. *Science.* 1991; 252(5007): 809-17.
12. Rebar E. J., Pabo C.O. Zinc finger phage: affinity selection of fingers with new DNA binding specificities. *Science.* 1994; 263(5147): 671-3.
13. Pabo C.O., Peisach E., Grant R.A. Design and selection of novel Cys2His2 zinc finger proteins. *Annu. Rev. Biochem.* 2001; 70: 313-40.
14. Li L., Wu L.P., Chandrasegaran S. Functional domains in Fok I restriction endonuclease. *Proc. Natl. Acad. Sci. USA.* 1992; 89(10): 4275-9.
15. Kim Y.G., Chandrasegaran S. Chimeric restriction endonuclease. *Proc. Natl. Acad. Sci. USA.* 1994; 91(3): 883-7.
16. Kim Y.G., Smith J., Durgesha M., et al. Chimeric restriction enzyme: Gal4 fusion to Fok I cleavage domain. *Biol. Chem.* 1998; 379(4-5): 489-95.
17. Bibikova M., Golic M., Golic K.G., et al. Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. *Genetics.* 2002; 161(3): 1169-75.
18. Bibikova M., Beumer K., Trautman J.K., et al. Enhancing gene targeting with designed zinc finger nucleases. *Science.* 2003; 300(5620): 764.
19. Townsend J.A., Wright D.A., Winfrey R.J., et al. High-frequency modification of plant genes using engineered zinc-finger nucleases. *Nature.* 2009; 459(7245): 442-5.
20. Osakabe K., Osakabe Y., Toki S. Site-directed mutagenesis in Arabidopsis using custom-designed zinc finger nucleases. *Proc. Natl. Acad. Sci. USA.* 2010; 107(26): 12034-9.

21. Curtin S.J., Zhang F., Sander J.D., et al. Targeted mutagenesis of duplicated genes in soybean with zinc-finger nucleases. *Plant Physiol.* 2011; 156(2): 466-73.
22. Beumer K., Bhattacharyya G., Bibikova M., et al. Efficient gene targeting in Drosophila with zinc finger nucleases. *Genetics.* 2006; 172(4): 2391-403.
23. Meng X., Noyes M.B., Zhu L.J., et al. Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nat. Biotechnol.* 2008; 26(6): 695-701.
24. Geurts A.M., Cost G.J., Freyvert Y., et al. Knockout rats via embryo microinjection of zinc-finger nucleases. *Science.* 2009; 325(5939): 433.
25. Carbery I.D., Ji D., Harrington A., et al. Targeted genome modification in mice using zinc-finger nucleases. *Genetics.* 2010; 186(2): 451-9.
26. Urnov F.D., Miller J.C., Lee Y.L., et al. Highly efficient en dogenous human gene correction using designed zinc-finger nucleases. *Nature.* 2005; 435(7042): 646-51.
27. Carroll D. Progress and prospects: zinc-finger nucleases as gene therapy agents. *Gene Ther.* 2008; 15(22): 1463-8.
28. Lee H.J., Kim E., Kim J.S. Targeted chromosomal deletions in human cells using zinc finger nucleases. *Genome Res.* 2010; 20(1): 81-9.
29. Sebastiano V., Maeder M.L., Angstman J.F., et al. *In situ* genetic correction of the sickle cell anemia mutation in human induced pluripotent stem cells using engineered zinc finger nucleases. *Stem Cells.* 2011; 29(11): 1717-26.
30. Tasan I., Jain S., Zhao H. Use of genome-editing tools to treat sickle cell disease. *Hum. Genet.* 2016; 135(9): 1011-28.
31. Li H., Haurigot V., Doyon Y., et al. *In vivo* genome editing restores haemostasis in a mouse model of haemophilia. *Nature.* 2011; 475(7355): 217-21.
32. Garriga-Canut M., Agustin-Pavon C., Herrmann F., et al. Synthetic zinc finger repressors reduce mutant huntingtin expression in the brain of R6/2 mice. *Proc. Natl. Acad. Sci. USA.* 2012; 109(45): 3136-45.
33. Kiskinis E., Sandoe J., Williams L., et al. Pathways disrupted in human ALS motor neurons indentified through genetic correction of mutant SOD1. *Cell Stem Cell.* 2014; 14(6): 781-95.
34. Soldner F., Laganiere J., Cheng A.W., et al. Generation of isogenic pluripotent stem cells differing exclusively at two early-onset Parkinson point mutations. *Cell.* 2011; 146(2): 318-31.
35. Ousterout D.G., Kabadi A.M., Thakore P.I., et al. Correction of dystrophin expression in cells from Duchenne muscular dystrophy patients through genomic excision of exon 51 by zinc finger nucleases. *Mol. Ther.* 2015; 23(3): 523-32.
36. Perez E.E., Wang J., Miller J.C., et al. Establishment of HIV_1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nat. Biotechnol.* 2008; 26(7): 808-16.
37. Tebas P., Stein D., Tang W.W., et al. Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. *N. Engl. J. Med.* 2014; 370(10): 901-10.
38. Moore M., Klug A., Choo Y. Improved DNA binding specificity from polyzinc finger peptides by using strings of two-finger units. *Proc. Natl. Acad. Sci. USA.* 2001; 98(4): 1437-41.
39. Segal D.J., Crotty J.W., Bhakta M.S., et al. Structure of Aart, a designed six-finger zinc finger peptide, bound to DNA. *J. Mol. Biol.* 2006; 363(2): 405-21.
40. Segal D.J., Beerli R.R., Blancafort P., et al. Evaluation of a modular strategy for the construction of novel polydactyl zinc finger DNA-binding proteins. *Biochemistry.* 2003; 42(7): 2137-48.
41. Liu Q., Segal D.J., Ghiara J.B., et al. Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. *Proc. Natl. Acad. Sci. USA.* 1997; 94(11): 5525-30.
42. Bae K.H., Kwon Y.D., Shin H.C., et al. Human zinc fingers as building blocks in the construction of artificial transcription factors. *Nat. Biotechnol.* 2003; 21(3): 275-80.

43. Jantz D., Amann B.T., Gatto J.G.J., et al. The design of functional DNA-binding proteins based on zinc finger domains. *Chem. Rev.* 2004; 104(2): 789-99.
44. Wu H., Yang W.P., Barbas C.F. Building zinc fingers by selection: toward a therapeutic application. *Proc. Natl. Acad. Sci. USA.* 1995; 92(2): 344-8.
45. Beerli R.R., Barbas C.F. Engineering polydactyl zinc finger transcription factors. *Nat. Biotechnol.* 2002; 20(2): 135-41.
46. Bhakta M.S., Segal D.J. The generation of zinc finger proteins by modular assembly. *Methods Mol. Biol.* 2010; 649: 3-30.
47. Gersbach C.A., Gaj T., Barbas C.F. Synthetic zinc finger proteins: the advent of targeted gene regulation and genome modification technologies. *Acc. Chem. Res.* 2014; 47(8): 2309-18.
48. Kolb A.F., Coates C.J., Kaminski J.M., et al. Site-directed genome modification: nucleic acid and protein modules for targeted integration and gene correction. *Trends Biotechnol.* 2005; 23(8): 399-406.
49. Camenisch T.D., Brilliant M.H., Segal D.J. Critical parameters for genome editing using zinc finger nucleases. *Mini Rev. Med. Chem.* 2008; 8(7): 669-76.
50. Gordley R.M., Gersbach C.A., Barbas C.F. Synthesis of programmable integrases. *Proc. Natl. Acad. Sci. USA.* 2009; 106(13): 5053-8.
51. Segal D.J., Dreier B., Beerli R.R., et al. Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences. *Proc. Natl. Acad. Sci. USA.* 1999; 96(6): 2758-63.
52. Dreier B., Beerli R.R., Segal D.J., et al. Development of zinc finger domains for recognition of the 5'-ANN-3' family of DNA sequences and their use in the construction of artificial transcription factors. *J. Biol. Chem.* 2001; 276(31): 29466-78.
53. Dreier B., Fuller R.P., Segal D.J., et al. Development of zinc finger domains for recognition of the 5'-CNN-3' family DNA sequences and their use in construction of artificial transcription factors. *J. Biol. Chem.* 2005; 280(42): 35588-97.
54. Maeder M.L., Thibodeau-Beganny S., Osiak A., et al. Rapid «open-source» engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol. Cell.* 2008; 31(2): 294-301.
55. Gupta A., Christensen R.A., Rayla A.L., et al. An optimized two-finger archive for ZFN-mediated gene targeting. *Nat. Methods.* 2011; 9(6): 588-90.
56. Szybalski W., Kim S.C., Hasan N., et al. Class-IIS restriction enzymes – a review. *Gene.* 1991; 100: 13-26.
57. Kovall R.A., Matthews B.W. Type II restriction endonucleases: structural, functional and evolutionary relationships. *Curr. Opin. Chem. Biol.* 1999; 3(5): 578-83.
58. Vanamee E.S., Santagata S., Aggarwal A.K. FokI requires two specific DNA sites for cleavage. *J. Mol. Biol.* 2001; 309(1): 69-78.
59. Wah D.A., Bitinaite J., Schlidkraut I., et al. Structure of the restriction endonuclease FokI has implications for DNA cleavage. *Proc. Natl. Acad. Sci. USA.* 1998; 95(18): 10564-9.
60. Halford S.E., Catto L.E., Pernstich C., et al. The reaction mechanism of FokI excludes the possibility of targeting zinc finger nucleases to unique DNA sites. *Biochem. Soc. Trans.* 2011; 39(2): 584-8.
61. Bitinaite J., Wah D.A., Aggarwal A.K., et al. FokI dimerization is required for DNA cleavage. *Proc. Natl. Acad. Sci. USA.* 1998; 95(18): 10570-5.
62. Sanders K.L., Catto L.E., Bellamy S.R., et al. Targeting individual subunits of the FokI restriction endonuclease to specific DNA strands. *Nucleic Acids Res.* 2009; 37(7): 2105-15.

63. Pingoud A., Wende W. Generation of novel nucleases with extended specificity by rational and combinatorial strategies. *Chembiochem*. 2011; 12(10): 1495-500.
64. Porteus M.H., Baltimore D. Chimeric nucleases stimulate gene targeting in human cells. *Science*. 2003; 300(5620): 763.
65. Doyon Y., Vo T.D., Mendel M.C., et al. Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. *Nat. Methods.* 2011; 8(1): 74-9.
66. Miller J.C., Holmes M.C., Wang J., et al. An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat. Biotechnol.* 2007; 25(7): 778-85.
67. Ramalingam S., Kandavelou K., Rajenderan R., et al. Creating designed zinc-finger nucleases with minimal cytotoxicity. *J. Mol. Biol.* 2011; 405(3): 630-41.
68. Szczepek M., Brondani V., Buchel J., et al. Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nat. Biotechnol.* 2007; 25(7): 786-93.
69. Cornu T.I., Thibodeau-Beganny S., Guhl E., et al. DNA-binding specificity is a major determinant of the activity and toxicity of zinc-finger nucleases. *Mol. Ther.* 2008; 16(2): 352-8.
70. Gupta A., Meng X., Zhu L.J., et al. Zinc finger protein-dependent and -independent contributions to the in vivo off-target activity of zinc finger nucleases. *Nucleic Acids Res.* 2011; 39(1): 381-92.
71. Radecke S., Radecke F., Cathomen T., et al. Zinc-finger nuclease-induced gene repair with oligodeoxynucleotides: wanted and unwanted target locus modifications. *Mol. Ther*. 2010; 18(4): 743-53.
72. Handel E.M., Cathomen T. Zinc-finger nuclease based genome surgery: it's all about specificity. *Curr. Gene Ther.* 2011; 11(1): 28-37.
73. Moore J.K., Haber J.E. Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in Saccharomyces cerevisiae. *Mol. Cell Biol.* 1996; 16(5): 2164-73.
74. Delacote F., Lopez B.S. Importance of the cell cycle phase for the choice of the appropriate DSB repair pathway, for genome stability maintenance: The trans-S double-strand break repair model. *Cell Cycle.* 2008; 7(1): 33-8.
75. Smithies O., Gregg R.G., Boggs S.S., et al. Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. *Nature.* 1985; 317(6034): 230-4.
76. Capecchi M.R. Altering the genome by homologous recombination. *Science.* 1989; 244(4910): 1288-92.
77. Carroll D. Genome engineering with targetable nucleases. *Annu. Rev. Biochem.* 2014; 83: 409-39.
78. Lee C.M., Cradick T.J., Fine E.J., et al. Nuclease target site selection for maximizing on-target activity and minimizing off-target effects in genome editing. *Mol. Ther*. 2016; 24(3): 475-87.
79. Sander J.D., Dahlborg E.J., Goodwin M.J., et al. Selection-free zinc finger-nuclease engineering by context-dependent assembly (CoDA). *Nat. Methods*. 2011; 8(1): 67-9.
80. Carlson D.F., Fahrenkrug S.C., Hackett P.B. Targeting DNA with fingers and TALENs. *Mol. Ther. Nucleic Acids.* 2012; 1(1): e3.
81. Pattanayak V., Guilinger J.P., Liu D.R. Determining the specificities of TALENs, Cas9, and other genome editing enzymes. *Methods Enzymol.* 2014; 546: 47-78.
82. Fine E.J., Cradick T.J., Bao G. Identification of offtarget cleavage sites of zinc finger nucleases and TAL effector nucleases using predictive models. *Methods Mol. Biol.* 2014; 1114: 371-83.
83. Hendel A., Fine E.J., Bao G., et al. Quantifying on and off-target genome editing. *Trend Biotechnol.* 2015; 33(2): 132-40.

84. Chandrasegaran S., Carroll D. Origins of programmable nucleases for genome engineering. *J. Mol. Biol.* 2016; 428(5 Pt. B): 963-89.
85. Cox D., Platt R.J., Zhang F. Therapeutic genome editing: prospects and challenges. *Nat. Med.* 2015; 21(2): 121-31.
86. Carroll D. Genome editing: progress and challenges for medical applications. *Genome Med.* 2016; 8(1): 120.
87. Yin H., Kanasty R.L., Eltoukhy A.A., et al. Non-viral vectors for gene-based therapy. *Nat. Rev. Genet.* 2014; 15(8): 541-55.
88. Just S., Büning H. Key to delivery: The (Epi-)genome editing vector toolbox. *Methods Mol. Biol.* 2018; 1767: 147-66.
89. Colella P., Ronzitti G., Mingozzi F. Emerging issues in AAV-mediated in vivo gene therapy. *Mol. Ther. Methods Clin. Dev.* 2018; 8: 87-104.
90. Rossidis A.C., Stratigis J.D., Chadwick A.C., et al. In utero CRISPRmediated therapeutic editing of metabolic genes. *Nat. Med*. 2018; 24(10): 1513-8.
91. Villiger L., Grisch-Chan H.M., Lindsay H., et al. Treatment of a metabolic liver disease by *in vivo* genome base editing in adult mice. *Nat. Med.* 2018; 24(10): 1519-25.
92. Yin C., Zhang T., Qu X., et al. In vivo excision of HIV-1 provirus by saCas9 and multiplex single-guide RNAs in animal models. *Mol. Ther.* 2017; 25(5): 1168-86.
93. Amoasii L., Hildyard J.C.W., Li H., et al. Gene editing restores dystrophin expression in a canine model of Duchenne muscular dystrophy. *Science.* 2018; 362(6410): 86-91.
94. Huang X., Zhou G., Wu W., et al. Genome editing abrogates angiogenesis *in vivo*. *Nat. Commun.* 2017; 8(1): 4-11.
95. Maeder M.L., Stefanidakis M., Wilson C.J., et al. Development of a gene-editing approach to restore vision loss in Leber congenital amaurosis type 10. *Nat. Med.* 2019; 25(2): 229-33.
96. Santiago-Fernandez O., Osorio F.G., Quesada V., et al. Development of a CRISPR/Cas9-based therapy for Hutchinson–Gilford progeria syndrome. *Nat. Med.* 2019; 25(3): 423-6.
97. Mingozzi F., High K.A. Therapeutic in vivo gene transfer for genetic disease using AAV: progress and challenges. *Nat. Rev. Genet.* 2011; 12(5): 341-55.
98. Mingozzi F., High K.A. Immune responses to AAV vectors: overcoming barriers to successful gene therapy. *Blood.* 2013; 122(1): 23-36.
99. Louis Jeune V., Joergensen J.A., Hajjar R.J., et al. Pre-existing anti-adenoassociated virus antibodies as a challenge in AAV gene therapy. *Hum. Gene Ther. Methods.* 2013; 24: 59-67.
100. Li L., Hu S., Chen X. Non-viral delivery systems for CRISPR/Cas9-based genome editing: challenges and opportunities. *Biomaterials.* 2018; 171: 207-18.
101. Glass Z., Lee M., Li Y., et al. Engineering the delivery system for CRISPR based genome editing. *Trends Biotechnol.* 2018; 36(2): 173-85.
102. Lee K., Conboy M., Park H.M., et al. Nanoparticle delivery of Cas9 ribonucleoprotein and donor DNA *in vivo* induces homology-directed DNA repair. *Nat. Biomed. Eng.* 2017; 1: 889-901.
103. Lee B., Lee K., Panda S., et al. Nanoparticle delivery of CRISPR into the brain rescues a mouse model of fragile X syndrome from exaggerated repetitive behaviours. *Nat. Biomed. Eng*. 2018; 2: 497-507.
104. Jang Y., Ye Z. Gene correction in patient-specific iPSCs for therapy development and disease modeling. *Hum. Genet.* 2016; 135: 1041-58.
105. Chal J., Pourquie O. Making muscle: skeletal myogenesis *in vivo* and *in vitro*. *Development*. 2017; 144(12): 2104-22.

106. Magli A., Perlingeiro R.R.C. Myogenic progenitor specification from pluripotent stem cells. *Semin. Cell Dev. Biol.* 2017; 72: 87-98.

107. van der Wal E., Herrero-Hernandez P., Wan R., et al. Large-Scale expansion of human iPSC-derived skeletal muscle cells for disease modeling and cell-based therapeutic strategies. *Stem Cell Reports.* 2018; 10(6): 1975-90.

108. Fiorotto R., Amenduni M., Mariotti V., et al. Liver diseases in the dish: iPSC and organoids as a new approach to modeling liver diseases. *Biochim. Biophys. Acta Mol. Basis Dis.* 2019; 1865(5): 920-8.

109. Hannoun Z., Steichen C., Dianat N., et al. The potential of induced pluripotent stem cell derived hepatocytes. *J. Hepatol.* 2016; 65(1): 182-99.

110. Brandão K.O., Tabel V.A., Atsma D.E., et al. Human pluripotent stem cell models of cardiac disease: from mechanisms to therapies. *Dis. Model. Mech.* 2017; 10(9): 1039-59.

111. Devalla H.D., Passier R. Cardiac differentiation of pluripotent stem cells and implications for modeling the heart in health and disease. *Sci. Transl. Med.* 2018; 10(435): eaah5457.

112. Genovese P., Schiroli G., Escobar G., et al. Targeted genome editing in human repopulating haematopoietic stem cells. *Nature*. 2014; 510(7504): 235-40.

113. Gaj T., Guo J., Kato Y., et al. Targeted gene knockout by direct delivery of zinc-finger nuclease proteins. *Nat. Methods*. 2012; 9(8): 805-7.

114. Liu J., Gaj T., Wallen M.C., et al. Improved cell-penetrating zinc-finger nuclease proteins for precision genome engineering. *Mol. Ther. Nucleic Acids*. 2015; 4(3): e232.

115. Loetscher P., Uguccioni M., Bordoli L., et al. CCR5 is characteristic of Th1 lymphocytes. *Nature*. 1998; 391(6665): 344-5.

116. Samson M., Libert F., Doranz B.J., et al. Resistance to HIV-1 infection in Caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature.* 1996; 382(6593): 722-5.

117. Huang Y., Paxton W.A., Wolinsky S.M., et al. The role of a mutant CCR5 allele in HIV-1 transmission and disease progression. *Nat. Med.* 1996; 2(11): 1240-3.

118. Li L., Krymskaya L., Wang J., et al. Genomic editing of the HIV-1 coreceptor CCR5 in adult hematopoietic stem and progenitor cells using zinc finger nucleases. *Mol. Ther.* 2013; 21(6): 1259-69.

119. DiGiusto D.L., Cannon P.M., Holmes M.C., et al. Preclinical development and qualification of ZFN-mediated CCR5 disruption in human hematopoietic stem/progenitor cells. *Mol. Ther. Methods Clin. Dev.* 2016; 3: 16067.

120. Liu N., Hargreaves V.V., Hong J., et al. Fetal hemoglobin (HbF) silencer BCL11A Acts through a novel DNA-motif in the gamma-globin promoters, simplifying the model for hemoglobin switching. *Blood*. 2017; 130(Suppl. 1): 283.

121. Xu J., Bauer D.E., Peng C., et al. Identification of BCL11A structure-function domains for fetal hemoglobin silencing. *Blood*. 2013; 122(21): 435.

122. Holmes M.C., Reik A., Rebar E.J., et al. A potential therapy for beta-thalassemia (ST-400) and sickle cell disease (BIVV003). *Blood*. 2017; 130(Suppl. 1): 2066.

123. Moran K., Ling H., Lessard S., et al. *Ex vivo* gene-edited cell therapy for sickle cell disease: disruption of the BCL11A erythroid enhancer with zinc finger nucleases increases fetal hemoglobin in plerixafor mobilized human CD34+ cells. *Blood*. 2018; 132(Suppl. 1): 2190.

124. Sharma R., Anguela X.M., Doyon Y., et al. *In vivo* genome editing of the albumin locus as a platform for protein replacement therapy. *Blood*. 2015; 126(15): 1777-84.

125. Ou L., DeKelver R.C., Rohde M., et al. ZFN-mediated *in vivo* genome editing corrects murine Hurler syndrome. *Mol. Ther.* 2019; 27(1): 178-87.

CHAPTER 2
TALE nucleases

Transcription activator-like effectors (TALEs) were first discovered in pathogenic *Xanthomonas* bacteria infecting many crop plants. It was found that functioning as eukaryotic transcription factors, these bacterial proteins participate in the regulation of gene expression in plants, affecting the behavior of plant cell genes [1–3]. In 2009, scientists deciphered the mechanism of sequence-specific DNA recognition by TALE [4, 5]. This milestone discovery triggered the generation of new TALEs recognizing specified DNA sequences. The fusion of the catalytic nuclease domain to TALEs gave birth to TALE nucleases or TALENs, a new tool for targeted DNA editing.

2.1. Discovery of TALEs

The study of TALEs started with the research in plant pathogenic bacteria. Having entered the nucleus of a plant cell, TALEs interact with the target sequence and activate genes of the host cell. Such interaction facilitates the spread of bacteria. Plants developed defense mechanisms against TALE pathogenic bacteria. Therefore, TALEs can function both and concurrently as virulence factors and avirulence factors recognizable by plants [6, 7]. Studies of *Xanthomonas campestris* pv. *Vesicatoria*, a pepper and tomato pathogen, resulted in the discovery of the *avrBs3* gene initiating the induction of hypersensitive response in plants having the *Bs3* resistance gene. The hypersensitive response is characterized by the rapid death of plant cells in the bacteria-affected region. As a result, the growth and spread of pathogens are localized and restricted, and the plant is protected from death.

The *avrBs3* gene was the first gene in the family of the so-called avirulence (*avr*) genes isolated later from different plant pathogenic bacteria and interacting in a gene-for-gene relation with dominant resistance genes [8]. The new *avr* genes found in pathogenic *Xanthomonas* lines associated with pathogenic symptoms and not related to dominant resistance genes of the respective plants triggered application of the *pathogenicity* (*pth*) term for some representatives of the family, including the

pthA gene [9]. Thus, representatives of *avrBs3*/*pthA*-like genes are generally classified as genes of this family.

Nucleotide sequences of this gene family have been found in different members of the genus *Xanthomonas*, for example, in *X. citri* infecting citrus plants [10], *X. oryzae* pv. *oryzae* and *X. campestris* pv. *Malvacearum* causing bacterial blight in rice [11] and cotton [12], respectively. The discovery of nuclear localization signals at the C-terminus of proteins encoded by the above genes led to the assumption of transcription involvement [13].

The rapidly increasing number of newly found effectors, most of them with unknown functions or phenotypes, called for a new term, first of all, required for avoiding confusion in the application of *avirulence* to plant and animal pathogens. The TALE term appeared due to effectors found in the C-terminal acidic activation domain (AAD) similar to the domain of eukaryotic transcription activators. The AAD is required for the avirulence activity of *AvrXa10* and *AvrXa7* genes [1, 14].

These properties of TALEs gave rise to the hypothesis of existing influence or interaction with the transcription apparatus of plant cells. One of the assumptions suggested binding to DNA; actually, it was found that the *AvrXa7* interacts with a double-stranded DNA within AT-rich regions [14]. However, no specific promoters or DNA binding sites were found in plant cells.

Another observation found *AvrBs3* and TALE-dependent transcripts in cycloheximide-resistant peppers [15]. It was believed that resistance to cycloheximide implies an immediate effect rather than a *de novo* synthesis of endogenous transcripts. At that time, nothing was known about specific TALE-dependent plant genes; therefore, there was no evidence of the existing cause and effect relationship with associated phenotypes. Later, it was shown that the interaction with the TFIIA transcription factor caused activation of gene expression in plants [16]. The direct proof of TALE binding to DNA was obtained only when *Bs3* and *Upa20* genes were found as targets for the *AvrBs3* [17, 18].

2.2. Structure of TALEs

Different protein structures can participate in binding proteins to specific DNA sequences. Generally, nucleoprotein interactions are governed by several amino acids, which jointly generate the specific recognition of a DNA sequence. Consequently, the protein-DNA binding specificity is extremely difficult to change. Conversely, the DNA binding specificity should be changed within a wide range to serve different tasks of biotechnology. Besides, the construct must be delivered to the cell nucleus during *in vivo* genome editing in eukaryotic cells. All these problems can be solved with TALE application. Structurally, TALEs can be divided into 3 domains (**Fig. 2.1**).

The N-terminal domain contains a type III secretion signal (T3, light-blue color) and is responsible for the translocation of *Xanthomonas spp* to a plant cell. The C-terminal domain harbors nuclear localization signals (NLS, dark-blue color) and a transcriptional acidic activation domain (AAD, orange color). The central DNA binding domain contains highly conserved tandem repeats of 34 amino acid residues

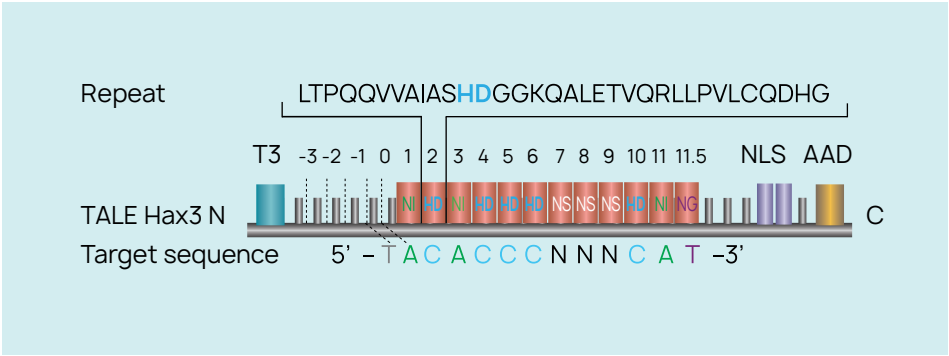


Fig. 2.1. Structure of natural TALE Hax3

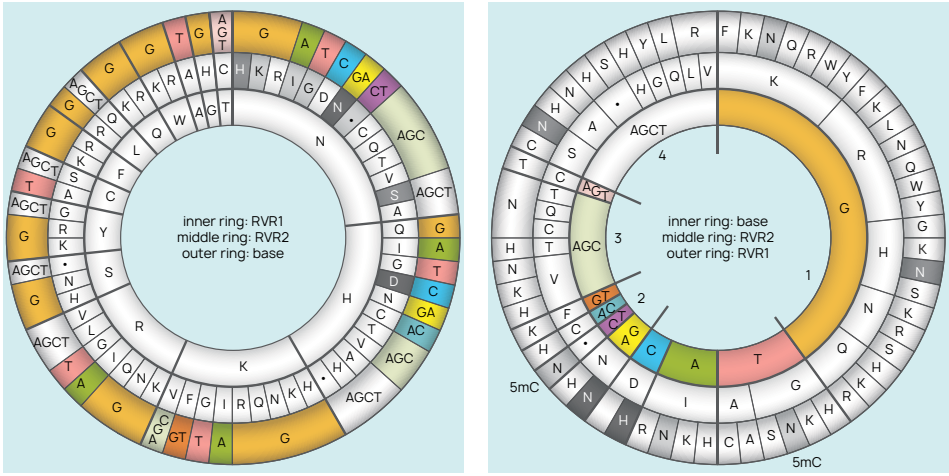
(red color). Each repeat recognizes one respective nucleotide in the coding strand of the target sequence; amino acids at positions 12 and 13 provide specific interaction. The additional thymine at the 5' end is recognized through the reduced repeat at position -1 in the N-terminal segment.

The N-terminal domain contains a type III secretion signal and is important for the initiation of DNA binding [19, 20]. The C-terminal domain contains nuclear localization signals and a transcriptional activation domain. The central domain is DNA binding and consists of repeats responsible for specific recognition of the target DNA sequence. Repeats are located tandem-wise, and each repeat comprises 33–35 amino acid residues (**Fig. 2.1**). The tandem repeat located close to the C-terminus contains only 20 amino acid residues and is referred to as a *half-repeat*. Among TALEs of the genus *Xanthomonas*, the central domain is highly conserved and is distinct in the number of repeats (from 1.5 to 30). In the repeated amino acid sequence, positions 12 and 13 are highly variable and are referred to as repeat variable diresidue (RVD). The number of repeats in TALEs determines the number of nucleotides in the TALE-recognized target sequence of DNA, and each RVD corresponds to a particular nucleotide [4, 5].

In most of the TALEs, the target sequence begins with a thymine nucleotide acting as an initiating base (T_0). In a number of cases, this rule may not work [21–23]. For example, for one of the TALE homologs of the *Ralstonia solanacearum* bacterium, the initiating base is guanine [24, 25], while cytosine is an initiating base for African *Xanthomonas oryzae* pv. *oryzae* (*Xoo*) BAI3 line [26]. Understanding the DNA sequence recognition mechanism opened opportunities for creating artificial TALEs with the required binding specificity.

2.3. RVD specificity

Today, there are 25 known naturally occurring RVDs; the most commonly used RVDs — HD (histidine, aspartic acid), NI (asparagine, isoleucine) and NG (asparagine, glycine) — are highly specific in recognizing cytosine, adenine and thymine, respectively (**Fig. 2.2**). Several RVDs are used for guanine: NH (asparagine, histi-



a *b*

Fig. 2.2. RVD specificity.
a — by a RVD; *b* — by a base

dine), NK (asparagine, lysine) and NN (asparagine, asparagine) recognizing guanine, and, in some cases, adenine [27]. Some RVDs can recognize more than 2 bases; for example, NS (asparagine, serine) and NA (asparagine, alanine) can be used almost as non-specific universal RVDs.

As of today, all the 400 theoretically possible RVD combinations have been analyzed [28]. In addition to the naturally occurring RVDs, a few new, functionally active RVDs have been discovered.

The activity comparable with the natural one has been demonstrated mostly by RVDs with histidine (H), lysine (K), asparagine (N) and arginine (R) at position 12. The specificity of the amino acid at position 13 correlates with the specificity of the known naturally occurring RVDs, for example, KI, NI, RI having specificity for adenine (**Fig. 2.2**).

The examination of the 3D structure of TALE-DNA complexes showed that each repeat consisted of two α-helices connected by a loop containing the RVD and reaching into the DNA [29, 30] (**Fig. 2.3**).

Only the amino acid at position 13 (Repeat Variable Residue 2, RVR2) interacts directly with the respective base from the sense strand and governs the RVD specificity (**Fig. 2.4**).

In contrast to RVR2, the amino acid at position 12 (RVR1) has an indirect impact on the DNA binding by stabilizing the RVD loop through its interaction with the amino acid at position 8 in the same repeat. Therefore, RVR1 has an indirect impact on the efficiency and specificity of RVR2 and DNA interaction.

In addition to the specific RVD-DNA interaction, 14–17 amino acid residues of each repeat interact with the sugar-phosphate backbone of the DNA sense strand.

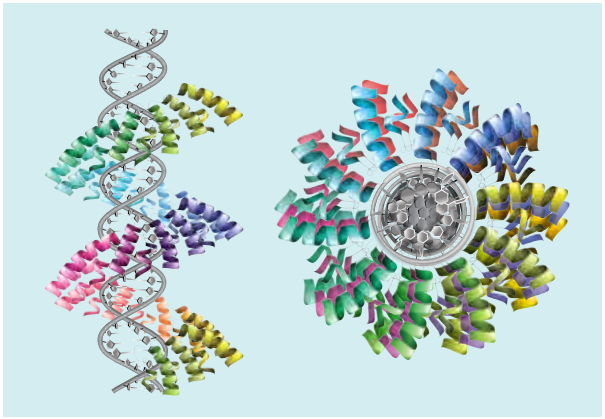


Fig. 2.3. Structure of the PthXo1 DNA binding region together with the target sequence

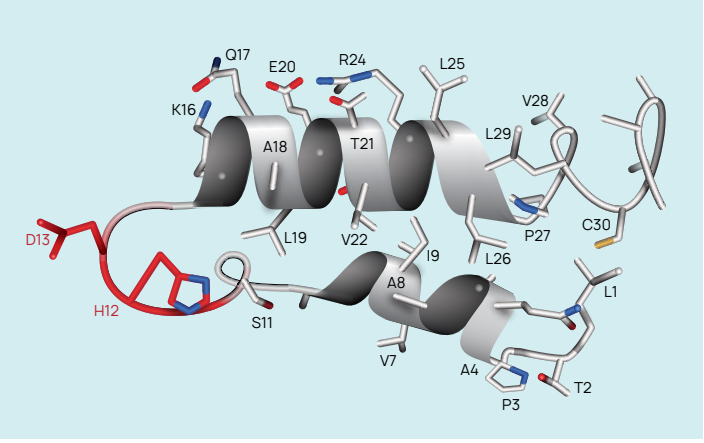


Fig. 2.4. Structure of a single PthXo1 repeat

The interaction between RVR2 and the base depends on 2 parameters: RVR2 binding to the base through hydrogen bonds or van der Waals interactions suggesting steric constraints for some RVR2+base combinations. When RVR2 is absent in some repeats, the star is used to denote the missing RVR residue, for example, N*. Because of steric constraints, such RVDs interact with pyrimidines in the DNA. In addition, N* can interact with 5-methylcytosine. The RVD NG can also interact with 5-methylcytosine. As a result, TALEs can be created for genomic regions with cytosine methylation when the standard cytosine-specific RVD (HD) cannot be used [31–34].

2.4. RVD efficiency

Studies of efficiency of different RVDs showed that there were *weak* and *strong* RVDs. TALEs composed of only weak RVDs did not affect the expression of target

genes either when the same RVD was repeated in the tandem or when different *weak* RVDs were used [27]. At least three *strong* RVDs (HD and NN) were required to restore the functional activity of TALEs. The RVD efficiency should meet the following parameters:

1. Similar specificity, but differences in RVR2; for example, such RVDs as NN, NH and NK recognize guanine with strong, intermediate and weak affinity, respectively, and the increase or decrease in the affinity strength can be 1,000-fold [35].

2. Similar specificity, but differences in RVR1: as opposed to the high variability of RVR2 amino acids, RVR1 is represented by a few different amino acids, mainly by asparagine (N) and histidine (H) [28, 36]. As RVR1 is connected with stability inside the repeat, it can indirectly influence the RVD and DNA interaction, i.e. RVDs possessing similar specificity can significantly differ in efficiency; for example, HD (histidine, aspartic acid) is stronger than ND (asparagine, aspartic acid) [27].

3. The RVDs recognizing different bases with different efficiency: For example, NN recognizes both purine bases, though preference is given to adenine rather than guanine, as the distribution of its negative charge is more suitable for interaction [37, 38]. Thus, NN is a strong RVD for guanine and a weak RVD for adenine [4, 27, 35]. Although NS is a universal RVD for all 4 bases, the RVR2 serine gives preference to purines (A, G) over pyrimidines (T, C).

Thus, RVD specificity and efficiency should be taken into consideration when artificial TALEs are created. Most of the created TALEs contain at least 3 strong RVDs [27, 39]. As the specificity can be varied for each position at the entry site, TAL effectors and respective TALENs offer more flexibility in genome editing than CRISPR/Cas systems.

2.5. Number of TALE repeats

Naturally occurring TALEs of the genus *Xanthomonas* contain 1.5–33.5 repeats; however, their functional activity is unknown [36]. To activate plant genes, artificial TALEs need at least 6.5 repeats; strong activation required min 10.5 repeats; the final minimum of repeats depends on the RVD composition [4]. For example, TALEs containing 10.5 repeats of only weak or intermediate RVDs (NI, NG, NH) do not trigger the expression of a target gene. The activity is restored when strong RVDs are added to the TALE structure or when the number of repeats is increased to 17.5 [27]. In some cases, when the TALE activity does not show any clear dependence on the number of repeats, the selection of an entry location for TALEs will influence the TALE functionality more strongly than the number of repeats [40]. It means that the TALE activity depends on the total contribution of each repeat to DNA binding, though only until the activity maximum is reached.

The fewer repeats the TALE structure has, the greater number of potential binding sites can be found in the genome, and the larger the genome is, the more sites can be found. For in the human genome, a fragment around 16 bp long must have a unique sequence. As high specificity of TALE binding is required for most of the biotechnical tools, when designing TALEs, the preference should be given to

the length that allows finding the unique sequence in the genome. For example, the length preferable for the human genome should be at least 20 bp or 18.5 repeats, considering the preceding thymine. In addition, TALEs should not contain RVDs specific for several bases. If there are RVDs with different binding efficiency, the priority should be given to the much stronger RVD; for example, NK should be used rather than NN for binding to guanine [41]. This approach to the TALE design provides its high specificity for genome editing.

2.6. Robustness against substitutions

Due to the ability of some RVDs to bind to several bases in the DNA strand, TALEs have a certain margin of strength towards substitutions in the target sequence. However, in some cases, several substitutions in the target sequence can completely impair the TALE functionality [42, 43]. The threshold, at which TALE binding is impaired crucially, may depend on the general efficiency of TALE binding (the RVD number and efficiency), the location of substitutions and the impact of individual substitutions on the 3D-structure of TALEs. A number of tests demonstrated that substitutions had a high impact on N-terminal repeats [35, 41, 44]. The smaller number of repeats makes TALEs less robust against substitutions [44]. The adverse impact on the total binding due to one substitution overweighs significantly the positive contribution of one matching repeat–base pair. Artificial TALEs with the central domain containing 17.5 repeats cannot recognize most of the targets with three substitutions or can activate the reporter gene only by 10% [44]. With three substitutions in the target sequence, artificial TALEs containing 13.5 repeats activate the reporter gene by reaching only 1% of the maximum level. TALEs containing 9.5 repeats cannot recognize the target sequence even with one substitution [44].

In contrast to the above said, robustness against substitutions in target sequence can be higher in TALEN pairs than in a single TALE derivative. The first TALEN monomer recognizing the target sequence can function as an anchor, while dimerization of nuclease domains will stabilize the bond of the second TALEN monomer, even if the target sequence match is far from perfect. The analysis of the off-target binding showed that TALENs containing 14.5 repeats retains functionality up to 6 substitutions in the target sequence [45]. Other studies showed that TALENs containing 17.5 repeats retained 30% of the activity with four substitutions [44]. In addition, TALENs with three substitutions at the 3'-end of the target sequence remained active compared to the respective TALEs [46]. It implies that binding of remote repeats is more significant for TALEs than for TALENs, possibly, because dimerization of nuclease domains does not need the complete binding of all repeats to DNA. All these examples demonstrate greater robustness against TALEN pair substitutions as compared to TALEs.

The recent study has revealed the impact of divalent metal cations (Mg^{2+} and Ca^{2+}) on the specificity of TALEs [47]. In the presence of these cations, the TALE non-specific binding rate decreased substantially due to the decreased binding rate of the N-terminal domain of TALEs, which was responsible for the non-specific interaction with DNA, and due to the concurrently increasing binding of the central

domain to the repeats responsible for the specific interaction with DNA. The effect was most pronounced for the TALEs that were 21.5 repeats long as compared to the shorter TALEs that were 11.5 and 15.5 repeats long.

2.7. Structural variability of repeats

Positions 12 and 13 are positions of amino acids most prone to variations in repeats; the other positions usually do not differ from each other in TALEs of the genus *Xanthomonas*. Yet, if we take a look at the variation of positions not included in RVDs, we will see that positions 4 and 24 are most variable [46] and most frequently are occupied by charged, polar or hydrophobic amino acids. For example, D (aspartic acid), E (glutamic acid) and A (alanine) — at position 4; R (arginine) and A (alanine) — at position 24. Negatively and positively charged amino acids generate electronegative or electropositive bands, respectively, on the opposite sides of spherical TALEs [29, 48]. In the study of the artificially constructed TALE with repeated RVDs (NN), each repeat contained charged aspartic acid (D) or polar glutamine (Q) at position 4; the TALE with aspartic acid demonstrated higher activity *in planta* [27]. Another study demonstrated the impact of amino acids at positions 4, 5 and 32 on the efficiency of binding to DNA [49].

In addition, the structure of repeats can vary in the number of amino acids. Most commonly, TALE repeats consist of 34 amino acids; however, repeats with 33 and 35 amino acids are not rare. When repeats contain 33 amino acids, it means that amino acid at position 13 is missing (for example, N*); when the repeat length is equal to 35 amino acids, it means that an additional residue, proline (P), is located after position 32 [36, 50]. Both length versions are functionally identical to repeats 34 amino acids long [4, 27].

Naturally occurring TALEs with repeats 30, 39, 40, and 42 amino acids long are relatively rare [36]. The above repeats have an interesting feature distinguishing them from standard repeats 33–35 amino acids long. A single long or short repeat in the array of repeats 34 amino acids long helps TALEs or TALENs recognize efficiently two different sequences: The first one is recognized in accordance with the general binding rule implying the one repeat — one base principle, and the other involving deletion of the base in the vicinity of the non-standard repeat [46]. Such flexibility is instrumental for creating TALEs and TALENs with double specificity for allelic versions with mutations of the insertion/deletion type.

2.8. Proteins of the TALE family

Besides *Xanthomonas spp*, proteins of the TALE family can be found in other species [51]. TALE-like gene activators in the pathogenic plant bacterium *Ralstonia solanacearum* contain a very similar region with repeats consisting of 35 amino acids, though the N and C-terminal domains differ from the domains of TALEs from the *Xanthomonas* family. *Ralstonia* TALE repeats differ from *Xanthomonas* repeats by some non-RVD amino acids and amino acids across repeats (**Fig. 2.5**) [24–27]. Despite the above differences, the specificity of *R. solanacearum* RVD TALE-proteins

is almost identical to the specificity of *Xanthomonas* RVD TALE-proteins in terms of recognizing DNA. The non-RVD amino acid variation may affect only the structure of the repeat array, base preference and RVD efficiency [24]. Proteins of the TALE family, which have repeats 33 amino acids long and very short N and C-terminal domains, have been found in symbiotic bacteria *Burkholderia rhizoxinica* dwelling within hyphae of *Rhizopus microspores* zygomycete fungi [51–53]. These bacteria are characterized by highly variable TALE repeats; differences can reach more than 50% [54], while similarities are observed only in some non-RVD amino acid positions in *Xanthomonas* or *Ralstonia*.

Nevertheless, the repeat region develops a TALE-like supercoiled structure, and repeats bind to DNA, using the same code as TALE-proteins in *Xanthomonas*. Thus, they are classified together with *Xanthomonas* and *Ralstonia* TALE-proteins as a family of proteins with the pronounced DNA binding ability.

TALEs can be used in the construction of specific nucleases and artificial gene activators, but they need the addition of the respective functional domains, as TALEs cannot activate genes [54–56]. It should be noted that repeats should not be used separately, as it may result in loss of the total activity [54], presumably, due to the fact that variability of non-RVD amino acid positions leads to significant interdependence of the architecture of neighboring repeats. Nevertheless, regardless of the differences among non-RVD positions in *Xanthomonas*, *Ralstonia,* and *Burkholderia* proteins of the TALE family, the crucial role in the DNA binding specificity belongs to the standard RVD code.

2.9. Construction of TALENs

To be inserted in the DNA, the double-strand break needs dimerization of *FokI* domains of a pair of TALENs, which are located on complementary strands and are oriented towards each other (**Fig. 2.5**). Such an arrangement tends to increase TALEN specificity in editing. The nuclease domain is typically located within 12–24 bp from the recognition site [57–59].

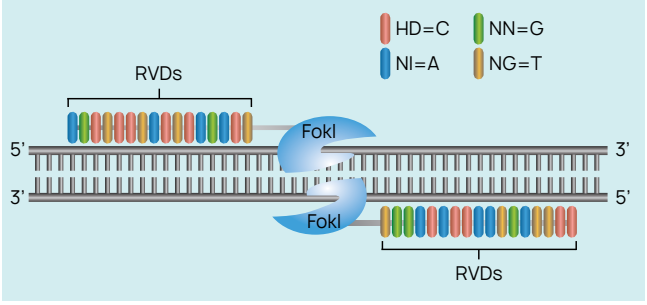


Fig. 2.5. TALEN structure and model.
The TALEN dimer: The DNA-binding domain consisting of an array of RVDs recognizing the specified sequence; the C-terminal domain is bound to the FokI endonuclease domain (blue color)

Some studies prove that the shorter C-terminal region of TALEs facilitated an increase in the nuclease activity of TALENs. Such shortening helps reduce the distance between two monomers, thus minimizing the risk of occurrence of non-specific cleavage. These studies also demonstrated that the removal of the nuclear localization domain does not affect nuclease activity [58, 59]. The catalytic domain of meganucleases can be used as an alternative nuclease domain, when designing TALENs, to increase their specificity [60]. There are also examples of the application of the TALEN monomer consisting of a TALE and modified I-SceI nuclease recognizing an 18-nucleotide sequence; the TALEN monomer of this type recognizes a 33-nucleotide-long β-globin gene sequence next to the mutation causing beta-thalassemia; in addition, the highly specific nuclease contributes to lower genotoxicity in cells [61].

Unspecified TALEs can be constructed within a few days with the help of plasmid libraries, the Golden Gate method of molecular cloning [62, 63], the ULtiMATE (**U**SER-based **L**iga**ti**on-**M**ediated **A**ssembly of **T**AL **E**ffector) system [64], chemical synthesis [65, 66] or ligation independent cloning [67].

2.10. Delivery of TALENs

TALEN monomers are readily delivered to the nucleus by DNA expression cassettes or as mRNA by using conventional transfection methods, such as microinjections or electroporation [68]. Although TALENs are big in size and have a large number of RVD repeats, their delivery with viral vectors can be very challenging. The adenovirus is used as a suitable vector for the delivery of constructs containing both TALEN monomers [69]. TALEN proteins can be delivered through covalent bonds of cell-penetrating transport peptides [70]. This method eliminates the risk of spontaneous integration of the TALEN-expressing DNA construct into the genome, which can take place during DNA delivery. This delivery method can also reduce the off-target nuclease activity due to the time-limited activity of injected TALENs [71].

2.11. Application of TALENs in science, biotechnology, and gene therapy

The library of TALENs targeted at protein-encoding human genes consists of 18,740 variants [72].

TALEN-based editing can be used for creating cellular models of human diseases, for studying mutations or genes causing the disease. This approach was used for creating genetic mutations associated with the disease, in models of somatic and stem cells designed for different human diseases [73], for example, beta-thalassemia [74]. Taking advantage of unsophisticated TALEN technology, researchers worked fast to create large genome deletions to study the functions of microRNA [75, 76]. High-performance TALEN complexes were also used in studies of a wide range of genes for epigenetic cancer-associated gene regulation; successful modifications of target genes accounted for more than 85% [65]. Thus, TALENs can be seen as a universal tool for studying both small and large genetic elements in complex genomes.

TALEs can be used in studies of epigenetic gene regulation, for example, cytosine methylation rates. It is possible due to the existence of methylcytosine [33] or carboxycytosine [77] specific RVDs making it possible to detect even single methylation sites [78].

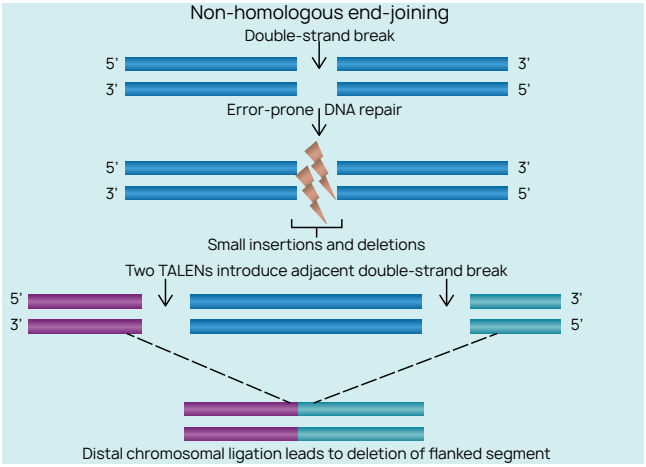
TALENs are used for efficient and fast modification of genes to create transgenic types or to knock out a specific gene (creation of knockout organisms). Nucleases induce a double-strand break in the target DNA, thus triggering DNA repair within a cell, and this repair can take either of two pathways: homologous recombination or NHEJ [79, 80] (**Fig. 2.6**).

During their studies involving TALENs and homologous recombination, researchers were able to insert a fluorescent reporter gene, which was 0.7–1.5 thousand bp in length and had germ-line transmission rates of 1.5–11.0% [81, 82]. It is a significant tool in studying functions of multiple genes in different organisms: the mouse [83–85], rat [86, 87], pig [88], cow [89], monkey [90], zebrafish [91–93], *O. latipes* [94], *C. elegans* [95, 96], newt [97], silk moth [98, 99], fruit fly [100, 101], mosquito [102, 103], frog [104], *T. gondii* [105]. Genome engineering has an unrivalled potential for solving problems in plant engineering [106, 107]. Numerous plant genes are arranged in arrays, which makes it difficult to modify selectively a single gene to study its functions. The ability of TALENs to detect relatively minor mismatches in nucleotide sequences turned them into a powerful tool for the modification of particular genes in arrays. Some of the examples are application of TALENs for fast modification of multiple genes in *Brachypodium* and rice; development of resistance to diseases in rice [108, 109], modification of genes in *Arabidopsis thaliana* [110] and barley [111]. In sugar cane studies, one pair of TALENs was enough to modify 107 out of 109 copies of the caffeic acid O-methyltransferase (*COMT*) gene responsible for lignin synthesis. The suppression of expression of this gene by using RNA interference resulted in improved production of bioethanol from lignocellulosic biomass. After being edited with TALENs, field plants demonstrated a 16.7% decrease in the production of lignin and a 43.8% increase in saccharification efficiency, while the biomass did not show any significant difference from the original type of the cane [112]. Therefore, TALENs can be safely used for the modification of plants characterized by high polyploidy levels. Modifications can be performed not only in nuclear but also in mitochondrial genes. For example, TALENs can be used for the removal of part of the gene sequence, associated with cytoplasmic male sterility, thus restoring the normal phenotype [113].

Application of artificial nucleases in gene therapy for the treatment of human genetic diseases has a promising future. Genome editing provides tools for direct correction of the mutations that caused the disease. For example, search for the treatment of muscular dystrophies resulted in the development of TALENs intended for creating small indels restoring the reading frame of the dystrophin gene [114, 115]. Besides, TALENs have been selected for the homologous recombination to correct mutations causing epidermolysis bullosa [45], sickle cell disease [116, 117], beta-thalassemia [118], xeroderma pigmentosum [119] and α_1-antitrypsin deficiency [120], as well as to correct the mitochondrial DNA by removing defective sequences [121].

TALENs are essential for the efficient introduction of mutations into the *CCR5* gene encoding one of the HIV co-receptors [58, 122–124]. Artificial TALENs have been constructed for the elimination of the hepatitis B virus genome from human cells [125, 126]. TALENs disrupting the myostatin gene were used to study the properties of the above gene [127]. The disruption of the myostatin gene leads to muscle hypertrophy. This approach can be used for the treatment of a number of

a



b

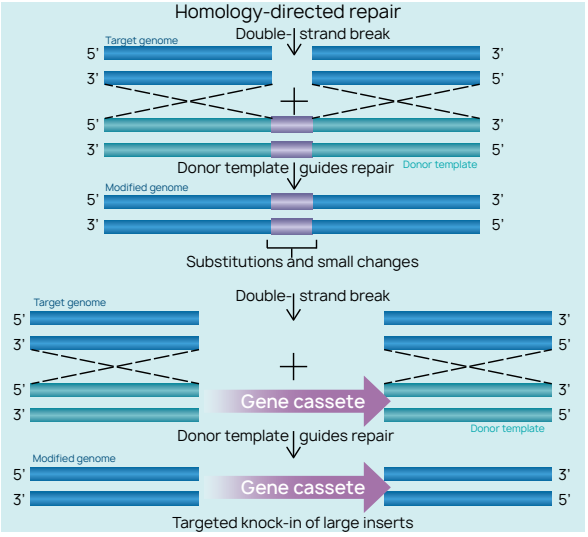


Fig. 2.6. Repair mechanism of a DNA double-strand break:
a — in the absence of a homologous DNA template, the break is repaired through NHEJ, which is prone to errors and can result in indels. However, this technique makes it possible to cut out the unwanted part of DNA from the genome.
b — in the presence of a DNA template, which is homologous to the target site, the region around the break goes through homologous repair. In this case, some modifications can be made in the DNA sequence or intact expressing constructs can be inserted in the required region of the genome

diseases, including muscular dystrophies. In studies on human cells, TALENs were used to create genomic translocations associated with malignant transformations [128], homozygous p53 mutant human embryonic stem cell lines [129]. E-7 targeted TALENs are used in editing papillomavirus oncogenes to cause the death of SiHa cells containing copies of integrated HPV 16 [130]. The authors see this approach as promising for the treatment of cervical cancer.

2.12. Online resources for the development of TALENs

There are quite a few programs designed for the selection and analysis of TALENs, many of them are available online. Below you can see some of them.

CHOPCHOP (URL: https://chopchop.cbu.uib.no) — a web-interface developed mostly by employees of the University of Bergen, Norway. Users can set the distance range between TALEN monomers, the size of TALENs, the number of mismatches to find off-target sites and can select RVDs for guanine (NN or NH). Users can also select specific oligonucleotides for target site amplification.

E-TALEN (URL: http://www.e-talen.org) — a web-interface of the German Cancer Research Center (Deutsches Krebsforschungszentrum) is designed for the selection of TALENs for knockout mutations, endogenous tagging, target excision repair. Users can select TALENs for 1 target or 50 target genes concurrently. Users can select *de novo* or evaluate the existing TALENs [131].

TALE-NT (URL: https://tale-nt.cac.cornell.edu) — a web-interface of the laboratory of Doctor Bogdanove of Cornell University, the author of one of the first articles about deciphering the TALE RVD code [5]. The website offers the following services: a selection of TALENs (registration of off-targets only for target sequences of max 1,000 bases), search for the best match for the selected RVD site sequence in DNA (Target Finder), search for the best match for paired RVDs (Paired Target Finder), search for potential non-specific entry sites for TALENs [132].

TALENoffer (http://www.jstacs.de/index.php/TALENoffer#TALENoffer_webserver) — a web-interface of the Martin Luther University of Halle-Wittenberg; it is designed for studying downloaded DNA sequences for off-target sites of the specified TALE.

Mojo Hand (URL: http://talendesign.org) — a web-interface of the Mayo Clinic network. The user can choose a manufacturer of restriction enzymes for target sequence sites.

PROGNOS (URL: http://bao.rice.edu/cgi-bin/prognos/prognos.cgi) — a web interface of the Biomolecular Engineering and Biomedicine Laboratory at Rice University. The web-based program is designed for processing of max 30,000 potential restriction sites; the program can be downloaded to a personal computer if a larger number of sites must be processed.

QueTAL (URL: http://bioinfo-web.mpl.ird.fr/cgi-bin2/quetal/quetal.cgi) — a web-interface designed for conducting a comparative functional and evolutionary analysis of different TALEs [133].

References

1. Zhu W., Yang B., Chittoor J.M., et al. AvrXA10 contains an acidic transcriptional activation domain in the functionally conserved C terminus. *Mol. Plant Microbe Interact.* 1998; 11(8): 824-32.
2. Schornack S., Meyer A., Romer P., Jordan T., Lahaye T.J. Gene-for-gene-mediated recognition of nuclear-targeted AvrBs3-like bacterial effector proteins. *Plant Physiol.* 2006; 163: 256-72.
3. Yang B., Sugio A., White F.F. Os8N3 is a host disease-susceptibility gene for bacterial blight of rice. *Proc. Natl. Acad. Sci. USA*. 2006; 103(27): 10503-8.
4. Boch J., Scholze H., Schornack S., et al. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*. 2009; 326(5959): 1509-12.
5. Moscou M.J., Bogdanove A.J. A simple cipher governs DNA recognition by TAL effectors. *Science*. 2009; 326(5959): 1501.
6. White F.F., Yang B., Johnson L.B. Prospects for understanding avirulence gene function. *Curr. Opin. Plant Biol*. 2000; 3(4): 291-8.
7. Bogdanove A.J., Schornack S., Lahaye T. TAL effectors: finding plant genes for disease and defense. *Curr. Opin. Plant Biol.* 2010; 13(4): 394-401.
8. Bonas U., Stall R.E., Staskawicz B.J. Genetic and structural characterization of the avirulence gene *avrBs3* from Xanthomonas campestris pv. vesicatoria. *Mol. Gen. Genet.* 1989; 218(1): 127-36.
9. Swarup S., De Feyter R., Brlansky R.H., et al. A pathogenicity locus from Xanthomonas citri enables strains from several pathovars of X. campestris to elicit cankerlike lesions on citrus. *Mol. Plant. Path.* 1991; 81(7): 802-9.
10. Swarup S., Yang Y., Kingsley M.T., et al. A Xanthomonas citri pathogenicity gene, pthA, pleiotropically encodes gratuitous avirulence on nonhosts. *Mol. Plant Microbe Interact.* 1992; 5: 204-13.
11. Hopkins C.M., White F.F., Choi S., et al. Identification of a family of avirulence genes from Xanthomonas oryzae pv. oryzae. *Mol. Plant Microbe Interact.* 1992; 5(6): 451-9.
12. De Feyter R., Yang Y., Gabriel D.W. Gene-for-genes interactions between cotton R genes and Xanthomonas campestris pv. Malvacearum avr genes. *Mol. Plant Microbe Interact.* 1993; 6(2): 225-37.
13. Yang Y., Gabriel D.W. Xanthomonas avirulence/pathogenicity gene family encodes functional plant nuclear targeting signals. *Mol. Plant Microbe Interact.* 1995; 8(4): 627-31.
14. Yang B., Zhu W., Johnson L.B., et al. The virulence factor AvrXa7 of Xanthomonas oryzae pv. oryzae is a type III secretion pathway dependent, nuclear-localized, double-stranded DNA binding protein. *Proc. Natl. Acad. Sci. USA*. 2000; 97(17): 9807-12.
15. Marois E., van den Ackerveken G., Bonas U. The Xanthomonas type III effector protein AvrBs3 modulates plant gene expression and induces cell hypertrophy in the susceptible host. *Mol. Plant. Microbe Interact.* 2002; 15(7): 637-46.
16. Iyer A.S., McCouch S.R. The rice bacterial blight resistance gene xa5 encodes a novel form of disease resistance. *Mol. Plant. Microbe Interact.* 2004; 17(12): 1348-54.
17. Kay S., Hahn S., Marois E., et al. A bacterial effector acts as a plant transcription factor and induces a cell size regulator. *Science*. 2007; 318(5850): 648-51.
18. Romer P., Hahn S., Jordan T., et al. Plant pathogen recognition mediated by promoter activation of the pepper Bs3 resistance gene. *Science*. 2007; 318(5850): 645-8.
19. Szurek B., Rossier O., Hause G., et al. Type III-dependent translocation of the Xanthomonas AvrBs3 protein into the plant cell. *Mol. Microbiol.* 2002; 46(1): 13-23.
20. Gao H., Wu X., Chai J., et al. Crystal structure of a TALE protein reveals an extended N-terminal DNA binding region. *Cell Res.* 2012; 22(12): 1716-20.

21. Lamb B.M., Mercer A.C., Barbas C.F. Directed evolution of the TALE N-terminal domain for recognition of all 5′ bases. *Nucleic Acids Res.* 2013; 41(21): 9779-85.
22. Tsuji S., Futaki S., Imanishi M. Creatinga TALE protein with unbiased 5′-T binding. *Biochem. Biophys. Res. Commun.* 2013; 441(1): 262-5.
23. Doyle E.L., Hummel A.W., Demorest Z.L., et al. TAL effector specificity for base 0 of the DNA target is altered in a complex, effector- and assay-dependent manner by substitutions for the tryptophan in cryptic repeat −1. *PLoS One.* 2013; 8(12): e82120.
24. de Lange O., Schreiber T., Schandry N., et al. Breaking the DNA-binding code of Ralstonia solanacearum TAL effectors provides new possibilities to generate plant resistance genes against bacterial wilt disease. *New Phytol*. 2013; 199(3): 773-86.
25. Li L., Atef A., Piatek A., et al. Characterization and DNA-binding specificities of Ralstonia TAL-like effectors. *Mol. Plant.* 2013; 6(4): 1318-30.
26. Schreiber T., Bonas U. Repeat 1 of TAL effectors affects target specificity for the base at position zero. *Nucleic Acids Res.* 2014; 42(11): 7160-9.
27. Streubel J., Blücher C., Landgraf A., et al. TAL effector RVD specificities and efficiencies. *Nat. Biotechnol.* 2012; 30(7): 593-5.
28. Yang J., Zhang Y., Yuan P., et al. Complete decoding of TAL effectors for DNA recognition. *Cell Res*. 2014; 24(5): 628-31.
29. Deng D., Yan C., Pan X., et al. Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science*. 2012; 335(6069): 720-3.
30. Mak A.N., Bradley P., Cernadas R.A., et al. The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science*. 2012; 335(6069): 716-9.
31. Valton J., Dupuy A., Daboussi F., et al. Overcoming transcription activator-like effector (TALE) DNA binding domain sensitivity to cytosine methylation. *J. Biol. Chem.* 2012; 287(46): 38427-32.
32. Deng D., Yin P., Yan C., et al. Recognition of methylated DNA by TAL effectors. *Cell Res.* 2012; 22(10): 1502-4.
33. Zhang Y., Liu L., Guo S., et al. Deciphering TAL effectors for 5-methylcytosine and 5-hydroxymethylcytosine recognition. *Nat. Commun*. 2017; 8(1).
34. Liu L., Zhang Y., Liu M., et al. Structural Insights into the Specific Recognition of 5-methylcytosine and 5-hydroxymethylcytosine by TAL Effectors. *J. Mol. Biol.* 2020; 432(4): 1035-47.
35. Meckler J.F., Bhakta M.S., Kim M.S., et al. Quantitative analysis of TALE-DNA interactions suggests polarity effects. *Nucleic Acids Res.* 2013; 41(7): 4118-28.
36. Boch J., Bonas U. Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annu. Rev. Phytopathol*. 2010; 48: 419-36.
37. Wicky B.I.M., Stenta M., Dal Peraro M. TAL effector specificity stems from negative discrimination. *PLoS One.* 2013; 8(11): e80261.
38. Deng D., Yan C., Wu J., et al. Revisiting the TALE repeat. *Protein Cell.* 2014; 5: 297-306.
39. Christian M.L., Demorest Z.L., Starker C.G., et al. Targeting G with TAL effectors: a comparison of activities of TALENs constructed with NN and NK repeat variable di-residues. *PLoS One.* 2012; 7(9): e45383.
40. Maeder M.L., Linder S.J., Reyon D., et al. Robust, synergistic regulation of human gene expression using TALE activators. *Nat. Methods*. 2013; 10(3): 243-6.
41. Garg A., Lohmueller J.J., Silver P.A., et al. Engineering synthetic TAL effectors with orthogonal target sites. *Nucleic Acids Res*. 2012; 40(15): 7584-95.

42. Römer P., Strauß T., Hahn S., et al. Recognition of AvrBs3-like proteins is mediated by specific binding to promoters of matching pepper Bs3 alleles. *Plant Physiol.* 2009; 150(4): 1697-712.
43. Strauß T., van Poecke R.M., Strauß A., et al. RNA-seq pinpoints a Xanthomonas TAL-effector activated resistance gene in a large-crop genome. *Proc. Natl. Acad. Sci. USA*. 2012; 109(47): 19480-5.
44. Mali P., Aach J., Stranges P.B., et al. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* 2013; 31(9): 833-8.
45. Osborn M.J., Starker C.G., McElroy A.N., et al. TALEN-based gene correction for epidermolysis bullosa. *Mol. Ther.* 2013; 21(6): 1151-9.
46. Richter A., Streubel J., Blücher C., et al. A TAL effector repeat architecture for frame shift binding. *Nat. Commun.* 2014; 5: 3447.
47. Cuculis L., Zhao C., Abil Z., et al. Divalent cations promote TALE DNA-binding specificity. *Nucleic Acids Res*. 2020; 48(3):1406-1422.
48. Stella S., Molina R., Yefimenko I., et al. Structure of the AvrBs3-DNA complex provides new insights into the initial thymine-recognition mechanism. *Acta Crystallogr. D Biol. Crystallogr*. 2013; 69(Pt. 9): 1707-16.
49. Tochio N., Umehara K., Uewaki J.I., et al. Non-RVD mutations that enhance the dynamics of the TAL repeat array along the superhelical axis improve TALEN genome editing efficacy. *Sci. Rep*. 2016; 6: 37887. doi: 10.1038/srep37887
50. Kay S., Boch J., Bonas U. Characterization of AvrBs3-like effectors from a *Brassicaceae* pathogen reveals virulence and avirulence activities and a protein with a novel repeat architecture. *Mol. Plant Microbe Interact.* 2005; 18(8): 838-48.
51. Schornack S., Moscou M.J., Ward E.R., et al. Engineering plant disease resistance based on TAL effectors. *Annu. Rev. Phytopathol.* 2013; 51: 383-406.
52. Partida-Martinez L.P., Hertweck C. Pathogenic fungus harbours endosymbiotic bacteria for toxin production. *Nature*. 2005; 437(7060): 884-8.
53. Lackner G., Moebius N., Partida-Martinez L., et al. Complete genome sequence of Burkholderia rhizoxinica, an Endosymbiont of Rhizopus microsporus. *J. Bacteriol.* 2011; 193(3): 783-4.
54. de Lange O., Wolf C., Dietze J., et al. Programmable DNA-binding proteins from Burkholderia provide a fresh perspective on the TALE-like repeat domain. *Nucleic Acids Res.* 2014; 42(11): 7436-49.
55. Juillerat A., Bertonati C., Dubois G., et al. BurrH: a new modular DNA binding protein for genome engineering. *Sci. Rep.* 2014; 4: 3831.
56. Stella S., Molina R., Lopez-Mendez B., et al. BuD, a helix-loop-helix DNA-binding domain for genome modifi cation. *Acta Crystallogr. D Biol. Crystallogr.* 2014; 70(Pt. 7): 2042-52.
57. Christian M., Cermak T., Doyle E.L., et al. Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics*. 2010; 186(2): 757-61.
58. Miller J.C., Tan S., Qiao G., et al. A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.* 2011; 29(2): 143-8.
59. Mussolino C., Morbitzer R., Lütge F., et al. A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res.* 2011; 39(21): 9283-93.
60. Boissel S., Jarjour J., Astrakhan A., et al. megaTALs: a rare-cleaving nuclease architecture for therapeutic genome engineering. *Nucleic Acids Res.* 2014; 42(4): 2591-601.
61. Lin J., Chen H., Luo L., et al. Creating a monomeric endonuclease TALE-I-SceI with high specificity and low genotoxicity in human cells. *Nucleic Acids Res*. 2015; 43(2): 1112-22.

62. Cermak T., Doyle E.L., Christian M., et al. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.* 2011; 39(12): e82.
63. Zhang F., Cong L., Lodato S., et al. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat. Biotechnol.* 2011; 29(2): 149-53.
64. Yang J., Yuan P., Wen D., et al. ULtiMATE system for rapid assembly of customized TAL effectors. *PLoS One.* 2013; 8(9): e75649.
65. Reyon D., Tsai S.Q., Khayter C., et al. FLASH assembly of TALENs for high-throughput genome editing. *Nat. Biotechnol.* 2012; 30(5): 460-5.
66. Briggs A.W., Rios X., Chari R., et al. Iterative capped assembly: rapid and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers. *Nucleic Acids Res.* 2012; 40(15): e117.
67. Schmid-Burgk J.L., Schmidt T., Kaiser V., et al. A ligation-independent cloning technique for high-throughput assembly of transcription activator-like effector genes. *Nat. Biotechnol.* 2013; 31(1): 76-81.
68. Wefers B., Meyer M., Ortiz O. et al. Direct production of mouse disease models by embryo microinjection of TALENs and oligodeoxynucleotides. *Proc. Natl. Acad. Sci. USA*. 2013; 110(10): 3782-87.
69. Holkers M., Maggio I., Liu J., et al. Differential integrity of TALE nuclease genes following adenoviral and lentiviral vector gene transfer into human cells. *Nucleic Acids Res.* 2013; 41(5): e63.
70. Liu J., Gaj T., Patterson J.T., et al. Cellpenetrating peptide-mediated delivery of TALEN proteins via bioconjugation for genome engineering. *PLoS One.* 2014; 9(1): e85755.
71. Gaj T., Guo J., Kato Y., et al. Targeted gene knockout by direct delivery of zinc-finger nuclease proteins. *Nat. Methods.* 2012; 9(8): 805-7.
72. Kim Y., Kweon J., Kim A., et al. A library of TAL effector nucleases spanning the human genome. *Nat. Biotechnol.* 2013; 31(3): 251-8.
73. Ding Q., Lee Y.K., Schaefer E.A., et al. A TALEN genome-editing system for generating human stem cell-based disease models. *Cell Stem.* Cell. 2013; 12(2): 238-51.
74. Alexeeva V., Aydin I.T., Schaniel C., Stranahan A.W., et al. A human H1-HBB11-GFP reporter embryonic stem cell line (WAe001-A-2) generated using TALEN-based genome editing. *Stem Cell Res*. 2020; 45: 101837.
75. Kim Y.K., Wee G., Park J., et al. TALEN-based knockout library for human microRNAs. *Nat. Struct. Mol. Biol.* 2013; 20(12): 1458-64.
76. Zhang Z., Xiang D., Heriyanto F., et al. Dissecting the roles of miR-302/367 cluster in cellular reprogramming using TALE-based repressor and TALEN. *Stem. Cell Reports*. 2013; 1(3): 218-25.
77. Maurer S., Giess M., Koch O., Summerer D. Interrogating key positions of size-reduced TALE repeats reveals a programmable sensor of 5-carboxylcytosine. *ACS Chem. Biol*. 2016; 11: 3294–9.
78. Kubik G., Schmidt M.J., Penner J.E., Summerer D. Programmable and highly resolved *in vitro* detection of 5-methylcytosine by TALEs. *Angew Chem. Int. Ed*. 2014; 53: 6002-06.
79. Rouet P., Smih F., Jasin M. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol. Cell Biol.* 1994; 14(12): 8096-106.
80. Rouet P., Smih F., Jasin M. Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proc. Natl. Acad. Sci. USA.* 1994; 91(13): 6064-8.
81. Zu Y., Tong X., Wang Z. et al. TALEN-mediated precise genome modification by homologous recombination in zebrafish. *Nat. Methods*. 2013; 10: 329–31.

82. Shin J., Chen J., Solnica-Krezel L. Efficient homologous recombination-mediated genome engineering in zebrafish using TALE nucleases. *Development*. 2014; 141(19): 3807-18.
83. Kato T., Miyata K., Sonobe M., et al. Production of *Sry* knockout mouse using TALEN via oocyte injection. *Sci. Rep.* 2013; 3: 3136.
84. Wang H., Hu Y.C., Markoulaki S., et al. TALEN-mediated editing of the mouse Y chromosome. *Nat. Biotechnol.* 2013; 31(6): 530-2.
85. Szczerkowska K.I., Petrezselyova S., Lindovsky J. et al. Myopia disease mouse models: a missense point mutation (S673G) and a protein-truncating mutation of the Zfp644 mimic human disease phenotype. *Cell. Biosci*. 2019; 9: 21.
86. Tesson L., Usal C., Menoret S., et al. Knockout rats generated by embryo microinjection of TALENs. *Nat. Biotechnol*. 2011; 29(8): 695-6.
87. Zhou X., Zhang R., Zou Z. et al. Hypoglycaemic effects of glimepiride in sulfonylurea receptor 1 deficient rat. *Br. J. Pharmacol*. 2019; 176(3): 478-90.
88. Montag J., Petersen B., Flögel A.K. et al. Successful knock-in of hypertrophic cardiomyopathy-mutation R723G into the MYH7 gene mimics HCM pathology in pigs. *Sci. Rep*. 2018; 8(1): 4786.
89. Su X., Wang S., Su G. et al. Production of microhomologous-mediated site-specific integrated LacS gene cow using TALENs. *Theriogenology*. 2018; 119: 282-8.
90. Chu C., Yang Z., Yang J., et al. Homologous recombination-mediated targeted integration in monkey embryos using TALE nucleases. *BMC Biotechnol*. 2019; 19(1): 7.
91. Bedell V.M., Wang Y., Campbell J.M., et al. In vivo genome editing using a highefficiency TALEN system. *Nature*. 2012; 491(7422): 114-8.
92. Kise R., Okasato R., Kano K., et al. Identification and biochemical characterization of a second zebrafish autotaxin gene. *J. Biochem*. 2019; 165(3): 269-275.
93. Völkel P., Bary A., Raby L., et al. Ezh1 arises from Ezh2 gene duplication but its function is not required for zebrafish development. *Sci. Rep.* 2019; 9(1): 4319.
94. Matsuzaki Y., Sakuma T., Yamamoto T., Saya H. Establishment of pten knockout medaka with transcription activator-like effector nucleases (TALENs) as a model of PTEN deficiency disease. *PLoS One*. 2017; 12(10): e0186878.
95. Wood A.J., Lo T.W., Zeitler B., et al. Targeted genome editing across species using ZFNs and TALENs. *Science*. 2011; 333(6040): 307.
96. Lo T.W., Pickle C.S., Lin S., et al. Precise and heritable genome editing in evolutionarily diverse nematodes using TALENs and CRISPR/Cas9 to engineer insertions and deletions. *Genetics*. 2013; 195(2): 331-48.
97. Hayashi T., Sakamoto K., Sakuma T., et al. Transcription activator-like effector nucleases efficiently disrupt the target gene in Iberian ribbed newts (Pleurodeles waltl), an experimental model animal for regeneration. *Dev. Growth Differ.* 2014; 56(1): 115-21.
98. Ma S., Zhang S., Wang F., et al. Highly efficient and specific genome editing in silkworm using custom TALENs. *PLoS One.* 2012; 7(9): e45035.
99. Xu J., Dong Q., Yu Y., et al. Mass spider silk production through targeted gene replacement in Bombyx mori. *Proc. Natl. Acad. Sci. USA*. 2018; 115(35):8757-8762.
100. Liu J., Li C., Yu Z., et al. Efficient and specific modifications of the *Drosophila* genome by means of an easy TALEN strategy. *J. Genet. Genomics.* 2012; 39(5): 209-15.
101. Beumer K.J., Trautman J.K., Christian M., et al. Comparing zinc finger nucleases and transcription activator-like effector nucleases for gene targeting in Drosophila. *G3 (Bethesda).* 2013; 3(10): 1717-25.
102. Aryan A., Myles K.M., Adelman Z.N. Targeted genome editing in Aedes aegypti using TALENs. *Methods*. 2014; 69(1): 38-45.

103. Smidler A.L., Terenzi O., Soichot J., et al. Targeted mutagenesis in the malaria mosquito using TALE nucleases. *PLoS One.* 2013; 8(8): e74511.

104. Lei Y., Guo X., Liu Y., et al. Efficient targeted gene disruption in Xenopus embryos using engineered transcription activator-like effector nucleases (TALENs). *Proc. Natl. Acad. Sci. USA*. 2012; 109(43): 17484-9.

105. Chen H., Guo Y., Qiu Y., et al. Efficient genome engineering of Toxoplasma gondii using the TALEN technique. *Parasit Vectors*. 2019; 12(1):112.

106. Mahfouz M.M., Li L., Shamimuzzaman M., et al. *De novo*-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. *Proc. Natl. Acad. Sci. USA*. 2011; 108(6): 2623-8.

107. Voytas D.F. Plant genome engineering with sequence-specific nucleases. *Annu. Rev. Plant. Biol.* 2013; 64: 327-50.

108. Shan Q., Wang Y., Chen K., et al. Rapid and efficient gene modification in rice and Brachypodium using TALENs. *Mol. Plant.* 2013; 6(4): 1365-8.

109. Han J., Xia Z., Liu P., et al. TALEN-based editing of TFIIAy5 changes rice response to Xanthomonas oryzae pv. Oryzae. *Sci. Rep*. 2020; 10(1): 2036.

110. Christian M., Qi Y., Zhang Y., et al. Targeted mutagenesis of Arabidopsis thaliana using engineered TAL effector nucleases. *G3 (Bethesda).* 2013; 3(10): 1697-705.

111. Wendt T., Holm P.B., Starker C.G., et al. TAL effector nucleases induce mutations at a pre-selected location in the genome of primary barley transformants. *Plant. Mol. Biol.* 2013; 83(3): 279-85.

112. Kannan B., Jung J.H., Moxley G.W., et al. TALEN-mediated targeted mutagenesis of more than 100 COMT copies/alleles in highly polyploid sugarcane improves saccharification efficiency without compromising biomass yield. *Plant. Biotechnol. J*. 2018; 16(4): 856-66.

113. Kazama T., Okuno M., Watari Y. et al. Curing cytoplasmic male sterility via TALEN-mediated mitochondrial genome editing. *Nat. Plants*. 2019; 5(7): 722-30.

114. Ousterout D.G., Perez-Pinera P., Thakore P.I., et al. Reading frame correction by targeted genome editing restores dystrophin expression in cells from Duchenne muscular dystrophy patients. *Mol. Ther.* 2013; 21(9): 1718-26.

115. Li H.L., Fujimoto N., Sasakawa N. et al. Precise correction of the dystrophin gene in duchenne muscular dystrophy patient induced pluripotent stem cells by TALEN and CRISPR-Cas9. *Stem Cell Reports*. 2015; 4(1): 143–54.

116. Sun N., Liang J., Abil Z., et al. Optimized TAL effector nucleases (TALENs) for use in treatment of sickle cell disease. *Mol. Biosyst.* 2012; 8(4): 1255-63.

117. Voit R.A., Hendel A., Pruett-Miller S.M., et al. Nuclease-mediated gene editing by homologous recombination of the human globin locus. *Nucleic Acids Res.* 2014; 42(2): 1365-78.

118. Ma N., Liao B., Zhang H., et al. Transcription activator-like effector nuclease (TALEN)-mediated gene correction in integration-free beta-thalassemia induced pluripotent stem cells. *J. Biol. Chem.* 2013; 288(48): 34671-9.

119. Dupuy A., Valton J., Leduc S., et al. Targeted gene therapy of xeroderma pigmentosum cells using meganuclease and TALEN. *PLoS One*. 2013; 8(11): e78678.

120. Choi S.M., Kim Y., Shim J.S., et al. Efficient drug screening and gene correction for treating liver disease using patient-specific stem cells. *Hepatology.* 2013; 57(6): 2458-68.

121. Bacman S.R., Williams S.L., Pinto M., et al. Specific elimination of mutant mitochondrial genomes in patient-derived cells by mitoTALENs. *Nat. Med.* 2013; 19(9): 1111-3.

122. Mussolino C., Alzubi J., Fine E.J., et al. TALENs facilitate targeted genome editing in human cells with high specificity and low cytotoxicity. *Nucleic Acids Res.* 2014; 42(10): 6762-73.

123. Ye L., Wang J., Beyer A.I., et al. Seamless modification of wild-type induced pluripotent stem cells to the natural CCR5Delta32 mutation confers resistance to HIV infection. *Proc. Natl. Acad. Sci. USA.* 2014; 111(26): 9591-6.

124. Jin L., Deng Y., He N., et al. Polyethylenimine-mediated CCR5 gene knockout using transcription activator-like effector nucleases. *J. Biomed. Nanotechnol*. 2018; 14(3): 546-52.

125. Bloom K., Ely A., Mussolino C., et al. Inactivation of hepatitis B virus replication in cultured cells and in vivo with engineered transcription activator-like effector nucleases. *Mol. Ther.* 2013; 21(10): 1889-97.

126. Chen J., Zhang W., Lin J., et al. An efficient antiviral strategy for targeting hepatitis B virus genome using transcription activator-like effector nucleases. *Mol. Ther.* 2014; 22(2): 303-11.

127. Xu L., Zhao P., Mariano A., et al. Targeted myostatin gene editing in multiple mammalian species directed by a single pair of TALE nucleases. *Mol. Ther. Nucleic Acids.* 2013; 2(7): e112.

128. Piganeau M., Ghezraoui H., De Cian A., et al. Cancer translocations in human cells induced by zinc finger and TALE nucleases. *Genome Res.* 2013; 23(7): 1182-93.

129. Zhou R., Xu A., Wang D., et al. A homozygous p53 R282W mutant human embryonic stem cell line generated using TALEN-mediated precise gene editing. *Stem Cell Res*. 2018; 27: 131-5.

130. Shankar S., Prasad D., Sanawar R., et al. TALEN based HPV-E7 editing triggers necrotic cell death in cervical cancer cells. *Sci. Rep*. 2017; 7(1): 5500.

131. Heigwer F., Kerr G., Walther N., et al. E-TALEN: a web tool to design TALENs for genome engineering. *Nucleic Acids Res.* 2013; 41(20): e190.

132. Doyle E.L., Booher N.J., Standage D.S., et al. TAL effector-nucleotide targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. *Nucleic Acids. Res.* 2012; 40(Web Server issue): W117-22.

133. Pérez-Quintero A.L., Lamy L., Gordon J.L., et al. QueTAL: a suite of tools to classify and compare TAL effectors functionally and phylogenetically. *Front. Plant. Sci.* 2015; 6: 545.

CHAPTER 3
CRISPR associated nucleases

CRISPR is an acronym for *Clustered Regularly Interspaced Short Palindromic Repeats*. This name refers to the unique organization of short palindromic, repeated DNA sequences found in most of the archaeal genomes and approximately 50% of the bacterial genomes. CRISPR sequences comprise the most widely distributed family of repeated sequences in prokaryotes. CRISPR/Cas systems are currently in the spotlight of active research in biology. In 2019, more than 5,000 research papers were published, addressing development, exploration, and application of CRISPR/Cas nucleases (**Fig. 3.1**).

The targeted CRISPR/Cas genome editing system consists of two functional components: a guide RNA and a Cas protein. The guide RNA, in its turn, contains

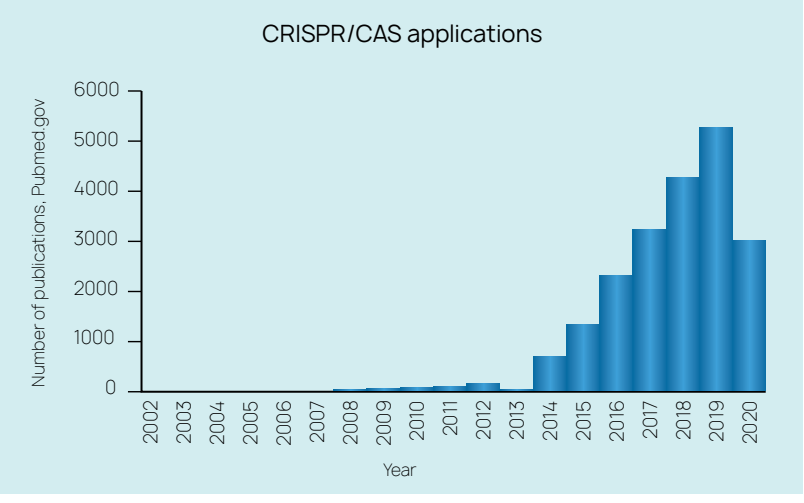


Fig. 3.1. Number of publications on CRISPR/Cas application.
Source: PubMed (URL: https://pubmed.ncbi.nlm.nih.gov)

a target-specific spacer and a conserved segment responsible for Cas protein binding. The Cas protein has nuclease activity. When the guide RNA and Cas protein form a sequence-specific ribonucleoprotein complex, the Cas protein cleaves the target DNA sequence.

3.1. CRISPR/Cas timeline

The first description of what would later be known as CRISPR came from Yoshizumi Ishino, a researcher at Osaka University, and his colleagues in 1987. They accidentally cloned part of a CRISPR sequence together with the iap gene, their target of interest, which is responsible for isozyme conversion of alkaline phosphatase in *Escherichia coli* [1], and discovered a peculiar arrangement of repeats — repeated sequences were interspersed with other sequences. The biological function of these unusual repeated sequences remained unclear till the mid-2000s.

In 1993, CRISPRs were for the first time found in *Haloferax mediterranei* archaea [2]; later, they were found in increasing numbers of bacterial and archaeal genomes. Francisco Mojica was the first researcher to characterize what is now called a CRISPR locus. He studied CRISPR loci throughout the 1990s, and in 2000, he recognized that what had been reported as disparate repeat sequences shared a common set of features:

- CRISPR loci are located in intergenic regions;
- they contain multiple short direct repeats with very little sequence variation;
- the repeats are interspersed with non-conserved sequences;
- the repeats have a common leader sequence, which consists of several hundred base pairs and is located on one side of the repeat cluster.

The analysis of sequenced bacterial and archaeal genomes uncovered the genomic context of CRISPR loci in many organisms and led to the discovery of four conserved genes regularly present in the vicinity of CRISPR regions. The genes were named CRISPR-associated (cas) genes (*cas1–cas4*) [3]. Cas1 and Cas2 did not display any similarity to functional domains of any known proteins, while Cas3 contained seven motifs typical of the superfamily 2 helicases, and Cas4 was found to be related to RecB exonucleases, which operate as an integral part of the RecBCD complex that repairs double-strand breaks during homologous recombination. Therefore, it was assumed that Cas3 and Cas4 participated in DNA metabolism, including DNA repair and recombination, transcriptional regulation and chromosome segregation. Their association with CRISPRs prompted the assumption that Cas proteins were involved in generating CRISPR loci [3].

While studying the *Streptococcus thermophilus* genome, Alexander Bolotin and his colleagues discovered an unusual CRISPR locus [4]. Although it was similar to the previously described systems, it lacked some of the known *cas* genes. Instead, the locus contained novel *cas* genes, including the gene encoding a large protein, which was assumed to have nuclease activity. Today, this protein is known as Cas9. Furthermore, the researchers noted that the spacers that were homologous

to viral genes share a common sequence at the same end. Later, it was found that this sequence, which was called PAM (protospacer adjacent motif), is required to recognize the target sequence.

At around the same time, K.S. Makarova and her colleagues conducted an independent systematic analysis of conserved gene contexts in all the prokaryotic genomes available in databases at that time and found several clusters of genes corresponding to *cas* genes (encoding hypothetical DNA polymerase, helicase, and RecB-like nuclease) in the genomes of hyperthermophilic archaea and two hyperthermophilic bacteria with available genome sequences, *Aquifex* and *Thermotoga* [5]. In the meantime, such conserved genes were not found in mesophilic and moderately thermophilic archaea and bacteria. This observation led to the assumption that these proteins could constitute a part of a DNA repair system specific to thermophilic organisms.

In 2005, F. Mojica noticed that sequences located between short repeats in CRISPR loci, the so-called CRISPR spacer regions or spacers, correspond to fragments from bacteriophage genomes [6]. At roughly the same time, other groups of scientists, working independently, published similar research findings [4, 6, 7]. Based on the literature review, F. Mojica and C. Pourcel pointed out that phages and plasmid DNAs do not infect host strains harboring homologous spacer sequences in CRISPR loci. Following on from the observations, they independently assumed that CRISPR sequences function within a biological defense system similar to the RNA interference system inherent in eukaryotic organisms, protecting cells from an invasion of foreign mobile genetic elements. It was also assumed that CRISPRs could somehow trigger a capture of fragments of the foreign DNA to create a *memory* of past genetic invasions [6, 7]. A. Bolotin and his colleagues confirmed these observations and pointed out the correlation between the number of spacers of phage origin and the level of resistance of the host strain to phage infection [4]. In their two publications in 2005, the authors acknowledged the previous discovery of *cas* genes and suggested that proteins encoded by these genes should be involved in the functioning of the new hypothetical prokaryotic immune system [4, 7].

In March 2006, E. Koonin studied clusters of orthologous groups of proteins by using computer analysis and offered a hypothetical scenario for CRISPR cascades acting as a bacterial immune system based on insertions homologous to phage DNA in a natural spacer array, renouncing the previous assumption that Cas proteins could incorporate a new DNA repair system [5].

The hypothesis asserting that the CRISPR/Cas system functions as an acquired prokaryotic immune system was finally proven in 2007, during experiments on the lactic acid bacterium *Streptococcus thermophiles* [8]. It was demonstrated that insertion of phage DNA sequences into a spacer region of *S. thermophilus* CRISPR made this strain resistant to the respective phage. On the other hand, the bacterial resistance to the phage infection disappeared when the respective sequence was deleted from the spacer region of *S. thermophiles* CRISPR. In addition, it was experimentally demonstrated that the CRISPR/Cas system was able to restrict the transformation of plasmids harboring sequences matching CRISPR spacers [9].

Soon after, scientists moved further to explore which CRISPR/Cas systems *stand in the way* of the invading phage. The first important information was received from John van der Oost and his colleagues who reconstituted the *immune system* by using the *E. coli* CRISPR discovered in 1987. They found that *E. coli* spacer sequences acquired from the phage are transcribed into small RNAs referred to as CRISPR RNAs (crRNA), which guide Cas proteins to the target DNA [10]. They also demonstrated that Cas9 was most likely the only protein required for inactivation of the invading phage; however, the underlying molecular mechanisms remained unknown.

Another key fact in understanding the mechanism of the CRISPR/Cas system was uncovered by L. Marraffini and E. Sontheimer, who demonstrated that DNA rather than RNA was the target molecule of the CRISPR/Cas9 system [9]. It was quite unexpected, as many scientists believed that the CRISPR/Cas system was in many ways similar to the eukaryotic RNA interference system, which is targeted at RNA. The authors pointed out that this system could become a powerful tool if it could be transferred to non-bacterial systems. However, looking ahead, we should admit that there can be another type of the CRISPR system to target RNA [11].

S. Moineau and his colleagues demonstrated that CRISPR/Cas9 generated double-strand breaks in the target DNA at precise locations, 3 nucleotides upstream PAM [12]. They also proved that Cas9 was the only protein required for cleavage in the CRISPR/Cas9 system. It is a distinctive characteristic of type II CRISPR systems, in which interference is mediated by a single large protein (Cas9 protein for CRISPR/Cas9) and a guide RNA.

The last piece of the puzzle of the CRISPR/Cas9 natural interference mechanism came from Emmanuelle Charpentier's group. They performed the sequencing of *Streptococcus pyogenes* RNA, which has a Cas9-containing CRISPR/Cas system, and found that in addition to crRNA there was a second short RNA, which they named trans-activating CRISPR RNA (tracrRNA) [13]. It turned out that tracrRNA formed a duplex with crRNA, and it was this duplex that directed Cas9 to its targets.

Soon after, it was demonstrated that the purified Cas9–CRISPR RNA (crRNA) complex can cleave the target DNA *in vitro* [14, 15]. Taking advantage of their heterologous system, V. Šikšnys and his colleagues purified Cas9-crRNA complex from the *E. coli* strain engineered to carry the *S. thermophiles* CRISPR locus and conducted a series of biochemical experiments to characterize the Cas9 mechanism [14]. They verified the cleavage site and requirements for PAM, and, using point mutations, showed that the RuvC domain cleaves the non-complementary strand, while the HNH domain cleaves the complementary site. They also pointed out that crRNA could be reduced to the size of 20 nucleotides, which was sufficient for efficient cleavage. The most impressive outcome was that they showed that Cas9 could be re-programmed by changing the crRNA sequence to target the desired site. The findings similar to those reported by Giedrius Gasiunas and his colleagues were reported almost at the same time by E. Charpentier in collaboration with Jennifer Doudna from the University of California, Berkeley [15]. They found that crRNA and tracrRNA could be fused to create a single synthetic guide RNA, further simplifying the system.

Then, the CRISPR/Cas system was used for genome editing in human nerve cells and mouse kidneys [16, 17]. F. Zhang and his team designed two different Cas9 orthologs (derived from *S. thermophilus* and *S. pyogenes*) and demonstrated targeted genome cleavage in human and murine cells [16]. P. Mali *et al* also showed that the system could be programmed to target multiple genomic loci and that it could guide a homology-directed repair [17].

Thus, CRISPR/Cas became well-known as a prokaryotic adaptive immune system [18, 19].

In 2013, the targeted genome editing CRISPR/Cas system was used to create the first modified genomes of plants, including rice, wheat, rockcress (*Arabidopsis*), tobacco, and sorghum [20–23]. In 2015, scientists, for the first time, edited genomes of human embryos, and in 2016, the first clinical trial of CRISPR/Cas9 was launched [24][1].

The discovery in 2017 was a milestone event in the CRISPR/CAS history when it was shown that CRISPR was a sensitive diagnostic tool for the detection of single DNA molecules or target RNAs. This high sensitivity has not been reported for other genome editing systems.

The CRISPR/Cas system is continuously being upgraded and modified. In 2019, scientists reported about a new DNA editing technique called *prime editing*, which offers high efficiency and precision in gene editing [25].

3.2. Proteins of the CRISPR/Cas system

Proteins of the CRISPR/Cas systems are characterized by a wide diversity; they are involved in different stages of CRISPR prokaryotic immunity; they have multiple predicted activities regarding nucleic acids, such as nuclease, helicase and polymerase activities [26–29].

Cas1 and Cas2 proteins are present in most of the known types of CRISPR/Cas systems and form a complex representing the adaptation module required for the integration of new spacers into the CRISPR arrays. At the expression stage, the CRISPR locus is transcribed, and the pre-crRNA transcript is processed into a mature crRNA by the type-specific Cas endonucleases. At the interference stage, crRNAs are bound to effector Cas endonucleases, and the corresponding complexes are recruited and cleave the target DNA or RNA in a sequence-dependent manner. Notably, unlike the adaptation module, Cas enzymes involved in the expression and interference stages vary from one CRISPR/Cas type to the other, and the same enzymes can participate in both stages of immunity.

The Cas1 protein, which works in complex with Cas2, captures the DNA of the invading bacteriophage. Cas1 acts as a molecular ruler, measuring out a fragment of DNA of the correct length. After capturing the invading bacteriophage DNA, the Cas1–Cas2 complex inserts this DNA fragment into the CRISPR array as a new spacer. Each spacer acts as a memory of a specific phage infection for future reference [30].

[1]URL: http://www.abedia.com/wiley/record_detail.php?ID=208

As mentioned previously, Cas2 works in complex with the Cas1 protein. Two copies of the Cas2 protein form a complex with 4 copies of Cas1. This complex searches the cell for invasive bacteriophage DNAs and/or plasmid DNAs [31].

Once the CRISPR/Cas system has identified the target – a double-stranded DNA, the complex is joined by the Cas3 protein, which, using the ATP energy, unwinds the invasive DNA of the attacking bacteriophage and destroys it. It is known that CRISPR/Cas3 can efficiently delete long DNA stretches from the target site in a human genome; such deletion was not as effective in traditional CRISPR/Cas9 systems [32].

The Cas4 protein is responsible for the recognition of correct PAMs and nucleotide motifs of protospacers; it defines the spacer length and its orientation. Cas4 allows the CRISPR/Cas system to integrate only new and functional spacers into the CRISPR array (a genomic locus containing several tandem CRISPRs) [33].

The Cas5 protein is bound to one end of the crRNA, blocking the complex terminus. In addition, Cas5 in complex with Cas6 processes or stabilizes pre-crRNA into separate crRNAs. Cas6 is required to generate crRNA in type I and III CRISPR/Cas systems [34].

The backbone of the CRISPR/Cas complex is formed by 6 copies of Cas7 protein. Each Cas7 monomer clams around crRNA and binds the target DNA [35].

In the CRISPR/Cas complex, Cas8 recognizes the invading bacteriophage DNA by identifying the PAM sequence; it helps unwind the target DNA and recruit Cas3 to destroy it [36].

Cas9 is a multi-domain protein having a DNA-binding and cleaving activity; it is guided by the crRNA-tracrRNA complex. tracrRNA is responsible for binding to the protein, while crRNA pairs with the complementary DNA of the target sequence. Once the target DNA has been bound, Cas9 induces a double-strand break into the target sequence.

Cas10 is a part of subtype IIIA CRISPR/Cas system *Staphylococcus epidermidis* and can be used as a tool for gene engineering of bacteriophages, including their design and construction [37].

Three Cas11 proteins form the backbone of the Cmr–Csm complex and guide the target RNA, thus facilitating the interaction between crRNA and the target DNA as well as the cleavage of the target DNA by subunits of the complex [38].

Cas12 (also known as Cpf1) is an RNA-guided DNA endonuclease. Unlike Cas9, after finding its target, Cas12 starts non-specific cleavage of a single-stranded DNA (ssDNA). For example, Cas12a can non-specifically cleave ssDNA after activation by the target molecule corresponding to its spacer sequence, thus becoming an attractive tool for detecting single copies of the target DNA in the solution [39].

Being an RNA-guided RNA endonuclease, Cas13 stands out among other Cas proteins. After the activation, Cas13 begins non-specifically to cleave the RNA. Similar to Cas12, Cas13 can be used in diagnosis [40].

Cas14 is a relatively small Cas protein, approximately twice as small as other Cas proteins in class 2 CRISPR systems. Unlike other Cas enzymes, Cas14 is found in archae a rather than in bacteria, thus leading to the assumption that Cas14 is a more primitive version of more complex Cas9 and Cas12 systems. Cas14 can bind

and cleave the target single-stranded DNA and, unlike Cas9, does not need PAM presence. In addition, Cas14 can be used for identification (diagnosis) of ssDNA [41].

3.3. Diversity and classification of CRISPR/Cas

It is amazing that closely related strains can vary significantly in their CRISPR content and distribution. For example, in mycobacteria (the *Mycobacterium* genus) CRISPR exists in *M. tuberculosis*, but not in *M. leprae*. On the other hand, phylogenetically distant *E. coli* and *Mycobacterium avium* as well as *Methanothermobacter thermautotrophicus* and *Archaeoglobus fulgidus* are characterized by almost identical CRISPR sequences [3, 42].

In reference to the CRISPR database[2], by June 2019, a total of 16,650 complete bacterial genomes and 340 complete archaeal genomes had been analyzed. CRISPR loci were found in 39.6% (6,595 out of 16,650) bacteria and in 79.4% (270 out of 340) archaea. Interestingly, 3.7% (608 out of 16,650) bacteria did not have any Cas proteins, while 3.7% (614 out of 16,650) bacterial genomes had Cas proteins, but did not have CRISPR arrays. At the same time, 4.1% (14 out of 340) archaea did not have Cas proteins, while 1.2% (four out of 340) had Cas protein, but did not have CRISPR arrays.

The number of CRISPR arrays in archaeal genomes can vary from one to 48, while the number of Cas protein clusters varies from one to six. At the same time, the number of CRISPR arrays in bacterial genomes varies from one to 121, and the number of Cas protein clusters — from one to nine. Multiple Cas protein clusters (from two to five) of the CRISPR/Cas system are significantly more frequently ($p < 0.01$) found in archaeal genomes[3].

Interestingly, the analysis of 1,724 genomes showed that CRISPR/Cas systems were much less prevalent in environmental microbial communities (10.4% in bacteria and 10.1% in archaea). The large difference between the prevalence estimated for complete genomes of cultivated microbes and uncultivated microorganisms is explained by the lack of CRISPR/Cas systems in major bacterial lineages that have no cultivated representatives [43].

The latest classification of CRISPR/Cas systems includes two classes based on encoded effector proteins [44]. Class 1 CRISPR/Cas systems work with multi-subunit effector complexes consisting of 4–7 Cas proteins present in an uneven stoichiometry. This system is widespread in bacteria and archaea, including all hyperthermophiles, and is found in about 90% of all identified CRISPR/Cas loci. The remaining 10% belong to class 2, use a single multi-domain effector protein and are found almost exclusively in bacteria [45].

Each class currently includes three types: Types I, III, and IV belong to class 1, and types II, V, and VI — to class 2. Types I, II, and III can be distinguished easily due to the presence of unique signature proteins: Cas3 for type I, Cas9 for type

[2]URL: http://crispr.i2bc.paris-saclay.fr

[3]URL: https://crisprcas.i2bc.paris-saclay.fr/MainDb/StrainList

II, and Cas10 for type III. Multi-subunit effector complexes of type I and type III systems, known as CRISPR-associated complexes for antiviral defense and Csm–Cmr complexes, respectively, are architecturally similar and evolutionary related [46–51]. Unlike other known CRISPR/Cas systems, the functionally uncharacterized type IV systems do not contain an adaption module consisting of Cas1 and Cas2 nucleases [46, 52]. Notably, effector modules of subtype IIIB systems use spacers created by type I systems, implying the modularity of CRISPR/Cas systems [53]. Although many of the genomes encoding type IV systems do not have identifiable CRISPR loci, it is not unlikely that type IV systems, which are similar to subtype IIIB systems, use crRNAs from different CRISPR arrays once these become available [52].

Each type of CRISPR/Cas systems, in its turn, is subdivided into several subtypes, and this further classification is based on additional signature genes and specific arrangements of these genes (type I is subdivided into A-E, F1-F3 subtypes, type III — into A-F subtypes, type IV — into A-C subtypes in class 1; type II — into A, B, C1-C2 subtypes, type V — into A, B1-B2, C-E, F1-F3, G, U1-U5 subtypes, type VI is subdivided into A, B1-B2, C, D subtypes in class 2) [44, 54]. Interestingly, CRISPR/Cas IF, IIA, IIB, IIC, VB, VIA, VIB1, VIB2, and VIC systems are found only in bacteria[4]. CRISPR/Cas IA, IB, ID, IIIB, IIIC, and IIID systems are significantly ($p < 0.01$) more frequently found in archaea, while IC and IE systems are more commonly found in bacteria.

3.4. CRISPR/Cas in targeted genome editing

The simple architecture of effector complexes has made class 2 CRISPR/Cas systems an attractive choice for developing a new generation of genome editing technologies. Several different class 2 effectors have been reported as suitable tools for targeted genome editing, including Cas9 in type II, Cas12a (formerly Cpf1), Cas12b (C2c1) in type V and Cas13a (C2c2) and Cas13b (C2c3) in type VI [44, 50].

3.4.1. Cas9

The most common and best understood multi-domain effector protein is Cas9, an RNA-dependent DNA endonuclease consisting of two unrelated nuclease domains — RuvC and HNH, which are responsible for inducing a double-strand break in the target DNA. Different types of guide RNA are used for specific targeting for the protein of the CRISPR/Cas9 system. For example, the guide RNA can represent a crRNA responsible for specific recognition of the target in complex with tracrRNA responsible for the binding of the enzyme of the CRISPR/Cas9 system and essential for pre-crRNA processing and target recognition in type II systems. In addition, the guide RNA can represent a single guide RNA (sgRNA), which combines characteristics of crRNA and tracrRNA in one molecule.

[4]URL: https://crisprcas.i2bc.paris-saclay.fr/MainDb/StrainList

The molecular mechanism of the target DNA cleavage by the Cas9–crRNA + tracrRNA complex is shown in **Fig. 3.2** [55].

Cas9 orthologs

Cas9 protein is found in 8.3% (1,375 out of 16,650) bacterial genomes. Currently, the Cas9 protein from the CRISPR/Cas9 *Streptococcus pyogenes* (SpCas9) system is the best described and the most popular molecular tool for genome editing. Shortly after Cas9 had become a popular tool for genome editing, scientists reported and characterized Cas9 orthologs from other bacteria: *Streptococcus thermophilus* (St1Cas9) and *Neisseria meningitidis* (NmCas9) [56], *Staphylococcus aureus* (SaCas9) [57], *Francisella novicida* (FnCas9) [58], *Campylobacter jejuni* (CjCas9) [59], *etc*. (**Table 3.1**).

Orthologs differ in their requirement for PAM, size (molecular weight) and specificity. Each ortholog binds a specific guide RNA, recognizing it by its unique component (*a hairpin*). Notably, different orthologs and their guide RNAs co-expressed in the same cell do not interfere with each other. This characteristic feature was initially used for multicolor visualization of genomic regions [60].

High-precision Cas9 variants

Non-specific activity of Cas9 was quite a serious problem, as Cas9 can edit the target DNA carrying up to five mismatches of its guide RNA [69–71]. This off-target effect was thoroughly analyzed *in vitro* and *in vivo* [72–74], and the CRISPR/Cas9 specificity criteria are as follows:

1) in most cases, the system cannot recognize the DNA site carrying more than 3 mismatches;

2) the CRISPR/Cas9 system cannot recognize and edit the DNA site with any number of mismatches in the proximity of PAM (within 10–12 bp);

3) the higher the CRISPR/Cas9 concentration is, the higher the likelihood of non-specific activity will be;

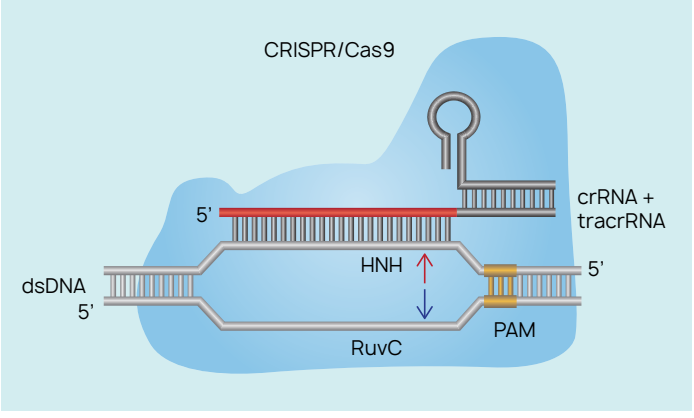


Fig. 3.2. Target DNA cleavage by the Cas9–crRNA + tracrRNA complex

Table 3.1. Cas9 *Streptococcus pyogenes* orthologs from other bacteria

No	CRISPR/Cas9 systems	PAM sequence	Reference
1	*Streptococcus pyogenes*	NGG	[61]
2	*Streptococcus thermophilus* CRISPR1	NNAGAAW NNAAAAW	[56, 62, 63]
3	*Streptococcus thermophilus* CRISPR3	NGGNG	[62]
4	*Streptococcus agalactiae*	NGG	[61]
5	*Listeria monocytogenes*	NGG	[61]
6	*Streptococcus mutans*	NGG	[64]
7	*Neisseria meningitidis*	NNNNGATT	[56, 65]
8	*Campylobacter jejuni*	NNNNACA	[63]
9	*Francisella novicida*	NG	[63]
10	*Streptococcus thermophilus* LMG18311	NNGYAAA NNNGYAAA	[66]
11	*Treponema denticola*	NAAAAN	[56]
12	*Staphylococcus aureus*	NNGRRT NNGRR(N)	[57]
13	*Enterococcus faecalis*	NGRNW	[67]
14	*Fusobacterium necrophorum*	NGG	
15	*Lactobacillus paracasei*	TNNAA	
16	*Lactobacillus rhamnosus*	NNNWR	
17	*Francisella tularensis subsp. novicida*	NGG YG	[58]
18	*Corynebacter diphtheriae*	NNRHHHY	[68]
19	*Sutterella wadsworthensis*	–	Bacterial Cas9 protein orthologs received at F. Zhang's laboratory (unpublished data)
20	*Legionella pneumophila str. Paris*	–	
21	*Filifactor alocis*	–	
22	*Staphylococcus pseudintermedius*	–	
23	*Lactobacillus johnsonii*	–	
24	*Streptococcus pasteurianus*	–	
25	*Lactobacillus farciminis*	–	
26	*Mycoplasma mobile*	–	
27	*Bacteroides coprophilus*	–	
28	*Fluviicola taffensis*	–	
29	*Flavobacterium columnare*	–	
30	*Sphaerochaeta globus str. Buddy*	–	
31	*Azospirillum B510*	–	
32	*Gluconacetobacter diazotrophicus*	–	
33	*Neisseria cinerea*	–	
34	*Roseburia intestinalis*	–	
35	*Parvibaculum lavamentivorans*	–	
36	*Nitratifractor salsuginis str DSM 16511*	–	
37	*Mycoplasma gallisepticum str. F*	–	
38	*Campylobacter lari CF89-12*	–	
39	*Pasteurella multocida*	–	

4) some 5'-NAG-3'-PAM sites can represent a target for the CRISPR/Cas9 system in bacteria and experiments *in vitro*; however, Cas9 has a much lesser affinity for NAG-PAM than for NGG-PAM.

In addition, the methods based on next-generation sequencing, such as GUIDE-seq [75], Digenome-seq [76], and ChIP-seq [77], can identify off-target sites for the CRISPR/Cas9 system. These high-throughput analytical methods confirmed that Cas9 has off-target activity and that the neat design of guide RNA is required to reduce the risk of non-specific activity of Cas9.

To increase the accuracy of targeted genome editing, Cas9 was modified into a Cas9 nickase (generating a single-strand break in the DNA strand) with D10A or H804A substitution and dCas9-FokI (catalytically inactive Cas9 protein fused with a FokI nuclease) [78–80]. These Cas9 nickases need a pair of guide RNAs to edit one site; therefore, they need a 40 bp sequence. Such length of the sequence reduces the likelihood of undesirable effects (non-specific editing). A fusion of Cas9 with additional DNA binding domains also reduces off-target editing [81]. Such modifications are indeed able to decrease the risk of non-specific editing. It should be noted that if the spacer sequence of the guide RNA is trimmed to 17–18 nucleotides, the precision of targeting improves respectively [82].

The problem of non-specific editing was solved with the help of an elegant approach. New Cas9 variants, which were more sensitive to mismatches, were created through the substitution of several amino acid residues in Cas9 [83, 69] (**Table 3.2**).

3.4.2. Cas9 nickases

Having introduced mutation to one of the two Cas9 nuclease domains, researchers created CRISPR nickases. The distinctive feature of nickases is that, unlike the Cas9 nuclease, they generate a single-strand break (a nick) in the DNA strand. It has been found that targeted editing coupled with using two guide RNAs and the Cas9 nickase (nCas9) reduces the likelihood of non-specific editing [78]. As single-strand breaks generally are quick to restore through homologous recombination and using of the undamaged complementary DNA strand as the template for repair, off-target effects of the Cas9 nickase are minimized. In the case of Cas9 from *Streptococcus pyogenes*, the D10A mutation inactivates the RuvC nuclease domain; therefore, this nickase cleaves only the target strand (complementary sgRNA). Alternatively, the H840A mutation in the HNH nuclease domain cleaves the off-target strand (**Fig. 3.3**) [89].

3.4.3. dCas9

In 2013, scientists conducted mutagenesis in catalytic nuclease domains of the Cas9 protein from *S. pyogenes* [90]. The study resulted in 2 introduced mutations: an H840A mutation in the HNH domain and a D10A mutation in the RuvC domain (**Fig. 3.4**). Thus, the catalytically inactive Cas9 protein, also called dCas9 or dCas9

Table 3.2. High-precision Cas9 variants

No	CRISPR/Cas9 systems	PAM sequence	Reference
1	Wild-type Cas9	NGG, NAG	[16, 84–86]
2	Cas9 with amino acid substitution D1135E (reduced binding to NAG)	NGG	[86]
3	Cas9 37R3-2 (the 37R3-2 intein integrated into Cas9) characterized by higher specificity	NGG	[87]
4	Cas9 with amino acid substitutions N497A-R661A-Q695A-Q926A; absence of significant off-target effects	NGG	–
5	Cas9 VRER variant	NGCG	[86]
6	Cas9 EQR variant	NGAG	[86]
7	Cas9 VQR variant	NGAN, NGNG	[86]
8	Cas9 with amino acid substitutions N497A-R661A-Q695A-Q926A, also known as Cas9-HF1; absence of significant off-target effects	NGG	[83]
9	Cas9 with amino acid substitutions K810A-K1003A-R1060A, also known as eSpCas9 (1.0); reduced number of off-target effects and robust cleavage of the target sequence	NGG	[88]
10	Cas9 with amino acid substitutions K848A-K1003A-R1060A, also known as eSpCas9 (1.1); absence of significant off-target effects	NGG	[88]

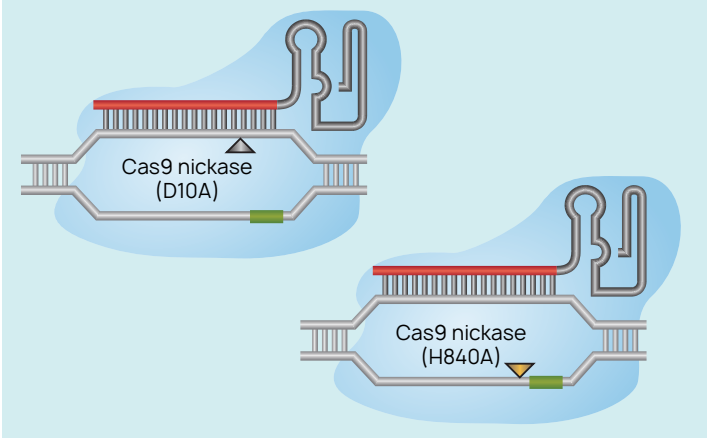


Fig. 3.3. Target DNA cleavage by the nCas9-sgRNA complex

null mutant, was received. Although dCas9 is not able to cleave the target DNA, it can still target and bind the intended DNA sequence with the same precision as the catalytically active Cas9 when guided by sgRNA. Unlike Cas9, instead of irreversibly

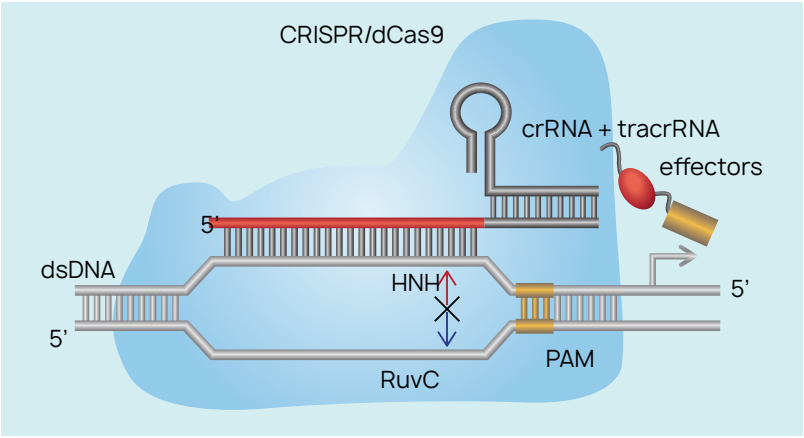


Fig. 3.4. Interaction between the target DNA and the dCas9–crRNA + tracrRNA complex

altering the genome, dCas9 interferes with the transcription of the target site, causing reversible *silencing* of the gene.

The application of dCas9 as a transcriptional activator (repressor) was only the beginning. Soon after, researchers began creating chimeric dCas9s with effector domains of repressor or activator proteins to use dCas9 targeting abilities for reversible gene activation, epigenomic editing and for many other purposes. Whether it is a promoter region, regulatory region or coding region, scientists can use dCas9 as a modular scaffold to facilitate effector attachment and enable the control of any gene without introducing irreversible DNA-damaging mutations.

Using the CRISPR/Cas9 system for gene activation

To endow dCas9 with gene activation abilities, dCas9 was first fused with classical transcriptional activators such as VP64 (a synthetic tetramer of the herpes simplex virus protein 16) or p65 (a transcription factor involved in many cellular processes). Although these systems demonstrated gene transcription activation across various eukaryotic cells, the level of activation was moderate (2–5-fold) [91].

To increase the activation level, a synergistic activation mediator (SAM) system was developed [92]. Being built on the basic dCas9-VP64 structure, this system also includes a guide RNA (sgRNA) modified to recruit additional transcriptional activators for a synergistic activation effect. The modified sgRNA contains two RNA hairpin aptamers that are able to bind to dimers of bacteriophage MS2 coat proteins. A fusion of MS2 proteins with additional activators such as p65 and the human heat shock factor 1 (HSF1) [91] results in the recruitment of 13 activation molecules per dCas9 molecule. This new dCas9–SAM system can reliably amplify gene expression from ten to multiple thousand-fold, depending on the baseline expression [91].

The dCas9–SAM system is an elegant and simple method, by using which scientists can selectively activate gene expression at a specific target within the

natural chromosomal context. The system is a tool essential in further fundamental research. Because of its activation ability, the system can become an unparalleled tool in the development of therapeutic interventions, genetic screening and transcriptional manipulations involving endogenous and synthetic genetic circuits across a variety of cell types [93, 94]. Researchers are already using the dCas9–SAM system in the activation of HIV-1 transcription to induce apoptosis and subsequent destruction of infected cells as well as in induction of transcription of latent HIV-1 proviral DNA for its complete elimination [95, 96].

It paved the way for designing another system combining 3 activators — VP64, p65, and RTA — the so-called dCas9-VPR system [97]. The dCas9-VPR system does not require an aptamer-modified sgRNA for effective activation and, therefore, its design process is much less complex. The gene activation achieved with this system was comparable with the dCas9-SAM system in the activation [94]. When combined with a sgRNA library, the dCas9-VPR system can be used in large-scale exploratory studies and functional screening assays, thus turning into a powerful tool for studying biological processes and pathways [92].

Modification of the CRISPR/Cas9 system for transcriptional repression

When dCas9 binds to the target site, its binding sterically interferes with the function of the transcriptional machinery, and this characteristic was used in the method known as CRISPR Interference or CRISPRi. CRISPRi can generate 1,000-fold transcriptional repression, efficiently knocking down gene expression in cells [93]. Although the system has demonstrated fairly good performance in bacteria, yeast, and other prokaryotic cells, it is less efficient in repressing gene expression in mammalian cells [98].

Attempts to overcome CRISPRi limitations resulted in the development of the dCas9-KRAB system, in which dCas9 is fused with KRAB (Krüppel-associated box) transcriptional repressor domain [98]. The system is designed to achieve transcriptional repression and is built on KRAB's ability to recruit a diverse array of histone modifiers that reversibly suppress gene expression by producing heterochromatin. This system made it possible to achieve a 60-80% reduction in the expression of highly specific endogenous eukaryotic genes during transient transfection [98]. Furthermore, the stably integrated dCas9–KRAB in HeLa caused a robust 5–10-fold repression of endogenous genes and promoter regions [98], with a 100-fold effect observed when the target site was 50–100 bp downstream of the transcription initiation site [90]. The presence of dCas9–KRAB did not have any effect on cell growth and viability [99].

Unlike other classical gene silencing approaches such as RNA interference (the method that knocks down gene expression through degradation of transcribed mRNA in the cytoplasm) [100], the dCas9–KRAB system offers reversible inhibition at the DNA level. This results in highly specific gene repression as well as in inhibition of non-coding RNAs, microRNAs, antisense transcripts, and nuclear-localized RNAs [98].

dCas9 mediated epigenetic editing

With the advent of genome engineering and editing technologies, we have been able to gain a better understanding of genes' ability to generate certain phenotypes. In addition to the genome, the phenotype can be affected by epigenetic modifications, which include modifications of the nucleosome and the DNA. Epigenetic regulation can affect the structure of a stretch of chromatin either by compressing it into a compact and transcriptionally inactive state (heterochromatin) or by opening it up for expression (euchromatin). Years of efforts in functional genomics have resulted in mapping and description profiles of millions of epigenetic regulatory elements in different tissues and cell types; however, current study methods focusing on each locus are labor-consuming, expensive and even toxic to live cells.

Chimeric dCas9 fused with different effector domains were created to study epigenomes at individual loci (**Table 3.3**). By using this epigenetics toolbox, scientists can study and manipulate gene regulation without altering the gene sequences. Such chimeric dCas9 systems can contribute to our understanding of the role of epigenetic changes in different molecular pathways and diseases, and it may offer therapeutic strategies for the treatment of such diseases.

dCas9-p300 — epigenetic activation through a modified CRISPR/Cas9 system

To gain a clearer idea about the mechanisms behind gene regulation, researchers need tools to modulate epigenetic tags with high specificity. Chimeric proteins of histone deacetylases or DNA methyltransferases fused to zinc-finger proteins or transcription activator-like effectors (TALEs) were created to modify the epigenome through targeted demethylation and deacetylation [101]. However, no chimeric proteins to acetylate histones have been created. In the meantime, histone acetylation is one of the most powerful systems for enhancing gene expression. I.B. Hilton *et al*. developed the dCas9-p300 system to solve the problem [102] through direct modification of the chromatin involved in a wide range of cellular pathways and processes.

In this system, dCas9 is fused to the catalytic domain of the human E1A-associated protein p300, the key component acetylating histones. The resulting system successfully activates gene expression when targeting either coding or regulatory regions, demonstrating its effectiveness as a transactivator of downstream genes [102]. The activation ranged from 50 to 10,000-fold when promoters or enhancers were

Table 3.3. Molecular tools of CRISPR/dCas9 for epigenetic modifications

Construct	Tool	Function	Gene expression
Histone modification	dCas9-p300	Acetylation	Activation
	dCas9-LSD1	Demethylation	Suppression
DNA methylation	dCas9-TET1CD	Demethylation	Activation
	dCas9-DNMT3A	Methylation	Suppression

targeted, and was highly specific judging by the transcriptome profiling [101, 102]. As this system employs mammalian p300, it also has minimum immunogenicity potential, thus being attractive for *in vivo* applications.

Thus, dCas9-p300 is a simple and unique tool for mapping out complex relationships between the epigenome, regulatory elements and the target gene's expression in functional genomics studies. By combining dCas9-p300 with inducible control, researchers will be able to activate genes in real-time and study the genome-wide activity of regulatory elements.

dCas9-LSD1 — epigenetic repression through a modified CRISPR/Cas9 system

The dCas9-LSD1 system is a gene repression system. In this system, dCas9 is fused to the lysine-specific histone demethylase 1 (LSD1) [103]. The dCas9-LSD1 system can repress downstream genes when targeting the distal enhancer region of the gene rather than its promoter [103], thus turning into a promising tool for studying regulatory activity enhancers.

As many genomic regions associated with human diseases are found within enhancer regions, the ability of dCas9–LSD1 to functionally map out enhancer elements in a highly specific manner makes it an indispensable tool in the exploration of enhancer-gene relationships. When used in combination with libraries of enhancer-specific guide RNAs, the dCas9-LSD1 system can provide a high-throughput and systemic way to identify all enhancers related to a gene.

dCas9-TET1CD — CRISPR/Cas9 system for targeted DNA demethylation

Targeted DNA methylation in mammalian cells generally occurs at the fifth carbon of cytosines within CpG dinucleotide sequences. Many processes, from cell development and differentiation to tumor genesis, can be regulated by DNA methylation; hypermethylation is known to be closely associated with cancer and neurological diseases [104]. The technology that can offer easy modulation of DNA methylation would open avenues for direct studying of functional relationships between the methylation status and gene expression.

The dCas9-TET1CD system is one of the technologies capable of editing the epigenome through targeted demethylation. In this system, the catalytically inactive Cas9 is fused to the catalytic domain (CD) of TET1 (Tet Methylcytosine Dioxygenase 1), an enzyme triggering DNA demethylation [104]. The guide RNA can be also modified to recruit MS2 bacteriophage coat proteins, which additionally carry two more TET1CD modules [105]. The system demonstrated its ability to generate a transcriptional increase in a gene array in a locus-specific manner, with minor off-target effects in different human and mouse cell lines [105].

The ability of dCas9-TET1CD to specifically and easily target the selected gene(s) will be helpful in the exploration of the functionality of DNA methylation in the regulation of gene expression in specific genomic contexts. The dCas9-TET1CD system was successfully used for epigenetic editing of the promoter of *BRCA1*, a tumor suppressor gene whose oversilencing through hypermethylation is associated

with breast and ovarian cancer [106]. This system can also be used to restore functional activity of other tumor suppressor genes essential for the fight against cancer and other diseases.

dCas9-DNMT3A — DNA methylation through a modified CRISPR/Cas9 system

Unlike histone-based control of cell phenotypes, DNA methylation is characterized by a more stable and long-term effect on gene expression. The dCas9-based methylation systems not only have the cross-species capability but also are not sensitive to CpG methylation. This makes them different from TALE-based systems, which cannot be used for epigenetic manipulations of mammalian promoters due to their CpG methylation sensitivity [107].

dCas9-DNMT3A is a methylating counterpart of the previously discussed dCas9-TET1CD system. This system involves a fusion of dCas9 - through a flexible glycine-serine linker (Gly_4Ser) — to the catalytic domain of DNMT3A, an active DNA methyltransferase, which can methylate CpG sites *in vivo*. It was found that the dCas9-DNMT3A system successfully induced site-specific CpG methylation distally and proximally from the promoter with the highest methylation activity (60%) observed 27 bp downstream of the PAM sequence. When several guide RNAs were involved, the effect of dCas9-DNMT3A was synergistic [108].

dCas9-DNMT3A has also been used for direct methylation of promoters of tumor suppressor genes whose hypermethylation is correlated with several kinds of cancer [109]. Thus, the dCas9-DNMT3A system, similar to other dCas9-based systems for epigenetic editing, can be used for functional genomics studies, epigenome editing and regulation.

Visualization of genomic loci with fluorophores

In addition to applications connected with epigenetic editing, catalytically inactive Cas9 fused to a fluorescent tag like green fluorescent protein (GFP) can be used for visualization of genomic loci in live cells and *in vivo*. The fluorescent labeling effect can be increased during visualization of target loci in the dCas9-GFP system with the help of guide RNAs with aptamers, which can recruit specific RNA-binding proteins labeled with fluorescent proteins. Comparing with such techniques as fluorescent hybridization *in situ* (FISH), CRISPR visualization offers a unique way of assessing chromatin dynamics in live cells.

CRISPR visualization can be multicolor and can offer concurrent tracking of several genomic loci in live cells and *in vivo*. One of the methods employs orthologous dCas9 (for example, *S. pyogenes* dCas9 and *S. aureus* dCas9) labelled with different fluorescent proteins.

Alternatively, guide RNAs can be fused to RNA aptamers specific to ortologous RNA-binding proteins tagged with different fluorescent proteins (CRISPRainbow) [110]. The CRISPR-Sirius tool intended for visualization of genomic loci offers eight different aptamers to modify guide RNAs [111] and to provide better stability and a stronger signal in the visualization of genomic loci.

The CRISPR visualization system can be used for tracking the dynamics of repetitive and non-repetitive genomic loci as well as for painting of chromosomes in live cells. Visualization of a specific genomic locus requires recruitment of many copies of labeled proteins to the given region. For example, chromosome-specific repetitive loci can be efficiently visualized in live cells by using a single guide RNA, which has multiple target sequences in the immediate vicinity. At the same time, the non-repetitive, unique genomic locus can also be labeled by co-delivering multiple guide RNAs that *overlap* the target locus. Chromosome painting requires the delivery of hundreds of guide RNAs with target sites throughout the chromosome [110, 112].

Isolation of target genomic regions with dCas9

Identification of molecules related to the specified genomic region *in vivo* is important for the understanding of locus functions. Using CRISPR/dCas9, researchers upgraded the chromatin immunoprecipitation (ChIP) technique to purify any genomic sequence the guide RNA can target [113–115].

In the enChIP (DNA-binding molecule-mediated chromatin immunoprecipitation) system, catalytically inactive Cas9 is used to purify genomic DNA bound by guide RNA. Epitope tag(s) intended for isolation can be fused to dCas9 or guide RNA. Various epitope tags, including 3×FLAG, PA and biotin tags, can be used for enChIP. In addition, Cas9-specific antibodies can be used to isolate target genomic regions with dCas9. The locus bound to dCas9 is then isolated by affinity purification against the epitope tag [113–118].

After the target genomic locus has been purified, all the locus-associated molecules can be identified by mass spectrometry (proteins), RNA sequencing (RNA) and NGS (other genomic regions) [115, 119–121]. Compared to traditional methods used for isolation of target genomic loci, CRISPR-mediated purification methods are less complex and suggest direct identification of molecules associated with the specified genomic region *in vivo*.

3.4.4. Base editing

Two classes of DNA base editors — cytosine base editors (CBEs) and adenine base editors (ABEs) — can be used to generate single-base changes in DNA without double-strand breaks (**Fig. 3.5**).

Cytosine base editors are created by fusing Cas9 nickase or catalytically inactive Cas9 to cytidine deaminase like APOBEC (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like). CBE editors are targeted at a specific DNA locus by using guide RNA, and they can convert cytidine to uridine within a small editing window near the PAM site. Uridine is subsequently converted to thymidine through base excision repair, creating a C to T substitution (or a G to A substitution on the complementary strand) [122–125].

Likewise, adenine base editors have been engineered to convert adenosine to inosine, which is treated like guanosine by the cell, creating an A to G (or T to C

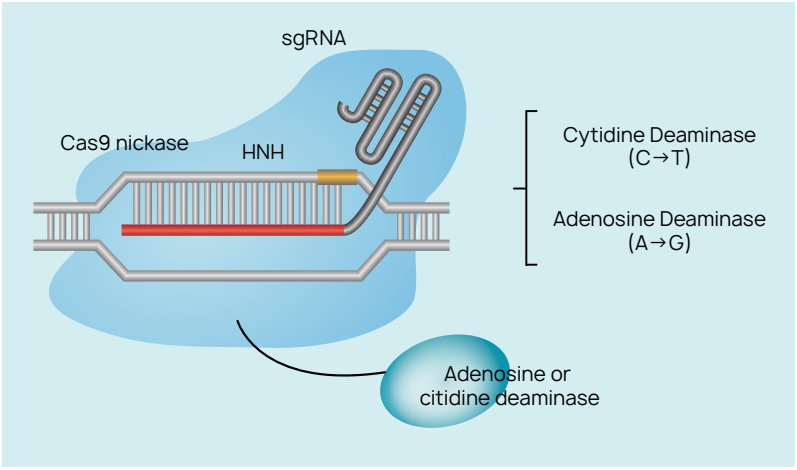


Fig. 3.5. nCas9-sgRNA-mediated base editing

on the complementary strand) change. Adenine DNA deaminases do not exist in nature, but have been created by directed evolution of the *Escherichia coli* TadA, a tRNA adenine deaminase. Like cytosine base editors, the evolved TadA domain is fused to a Cas9 protein to create the adenine base editor [126–129].

Both types of base editors are available with multiple Cas9 variants including modified Cas9 variants. Further advancements have been made by optimizing the expression of fusion proteins, by modifying the linker region between the Cas9 protein and deaminase to adjust the editing window or adding fusions that increase product purity such as the DNA glycosylase inhibitor (UGI) or the bacteriophage Mu-derived Gam protein (Mu-GAM) [130, 131].

While many base editors are designed to work in a very narrow window proximal to the PAM sequence, some base editing systems create a wide spectrum of single-nucleotide variants (somatic hypermutation) in a wider editing window, and, therefore, are well-suited to directed evolution applications [132].

3.4.5. Cas9 for prime editing

In October 2019, Andrew Anzalone *et al*. introduced a new genome-editing technique called prime editing [25] — a targeted editing method that can perform accurate targeted small insertions, deletions and base substitutions in the edited DNA sequence without double-strand breaks. Moreover, targeted sequence insertions can be achieved without the need for donor DNA templates. In addition, prime editing expands the limited scope of current DNA base editing abilities [133].

Similar to traditional genome editing, prime editing requires the presence of a Cas endonuclease and a single guide RNA. Instead of Cas9, this method uses Cas9 nickase — a variant of Cas9 that nicks only one strand of DNA rather than generating double-strand breaks. The Cas9 nickase is fused to a reverse transcriptase [25].

Currently, there are three variants of fusion proteins used for prime editing. The first version of fusion proteins was characterized by moderate editing efficiency; the second fusion protein contained further modifications that led to improved binding to target DNA and thermostability. The most recent versions of fusion proteins for prime editing include the ability to mend the mismatch sequences that occur with prime editing [25].

The guide RNA used in prime editing is substantially larger than the standard guide RNAs commonly used for CRISPR/Cas gene editing. This RNA is a guide RNA with a primer binding site (PBS) and a donor template containing the desired sequence added at the 3' end [25]. At present, such guide RNAs are created by using plasmid DNAs and *in vitro* transcription.

During the prime editing, the guide RNA complex binds to the target DNA, and Cas9 nicks only one strand, generating a single-strand break. The PBS, homologous to the target DNA and located on the guide RNA, binds to the target DNA fragment, and the donor RNA template is reverse transcribed (by using the reverse transcriptase incorporated into the fusion protein for prime editing) [25]. The target DNA is repaired with the new reversely transcribed DNA; the original DNA segment is removed by a cellular endonuclease. This leaves one strand of DNA edited and the other strand — unedited.

Third generation fusion or chimeric proteins are able to correct the unedited DNA strand by using an additional standard guide RNA. In this case, the unedited DNA strand is nicked by the Cas9 nickase, and the newly edited strand is used as a template to repair the nick, thus completing the editing [25].

While expanding the genome-editing toolbox is useful, the actual question is whether all these new tools can be used in therapeutic applications. A. Anzalone and co-workers demonstrated the ability of prime editing to model and correct common genetic diseases [25]. To prove the concept, they chose two diseases: sickle cell disease and Tay-Sachs disease. By using prime editing, the researchers created cell lines carrying mutations responsible for the above diseases. Then, they used the prime editor and guide RNAs harboring the wild type donor sequences of both genetic diseases to efficiently correct the mutations. Importantly, these studies show the vast opportunities opened by prime editing application for correcting single nucleotide mutations (sickle cell disease) and for longer nucleotide corrections (Tay-Sachs disease requiring a 4-bp insertion) [25].

As with all technological advancements, additional studies need to be performed to understand how prime editing works. Researchers continue to optimize the technique and try to find out if prime editing can be used in diverse cell types (especially therapeutically significant cell types such as primary and stem cell), what the long-term effects are (if any), and the extent of off-target effects.

3.4.6. Cas12

The gene initially called *cpf1* is found in several bacterial and archaeal genomes, typically in the same locus with *cas1* and *cas2* genes and a CRISPR array [44]. Cas12a (Cpf1), the prototype of type V multi-domain Cas protein ef-

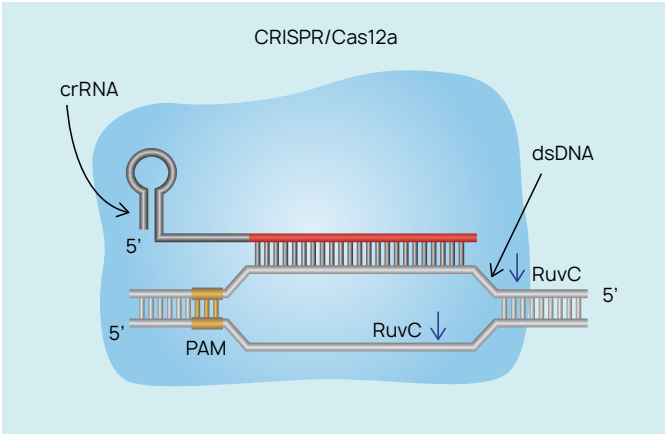


Fig. 3.6. Target DNA and Cas12a-crRNA interaction

fectors, contains two RuvC-like nuclease domains, but lacks an HNH domain. The structural analysis of the Cas12a–crRNA-target DNA complex revealed a second nuclease domain (Nuc) with a unique structure functionally identical to the HNH domain of Cas9 [134] (**Fig. 3.6**). Cas12a is a single-RNA-guided endonuclease that does not need a tracrRNA, which is essential for Cas9 activity [135]. The protein also differs from Cas9 in its cleavage pattern and PAM recognition, which identifies target strands. The variety of Cas12 proteins is shown in **Table 3.4**.

3.4.7. Cas13

The discovery of two distantly related class 2 effector proteins, Cas9 and Cas12a, suggested that there could be other variants of such systems. Indeed, soon after, the directed bioinformatics search for class 2 effectors uncovered Cas12b (type V), Cas13a and Cas13b (type VI) proteins, which were different from Cas9 and Cas12a; their activity was confirmed [135]. Type V effectors, similar to Cas9, need a tracrRNA for the targeted activity. Most of the functionally characterized CRISPR/Cas systems have been reported to target DNA, and only multicomponent type IIIA and IIIB systems can cleave RNA [142]. Type VI effectors Cas13a and Cas13b specifically target RNA and, therefore, mediate RNA interference. Unlike type II and type V effectors, Cas13a and Cas13b lack characteristic RuvC-like nuclease domains and instead contain a pair of HEPN (higher eukaryote and prokaryote nucleotide) binding domains [143] (**Fig. 3.7**). The diversity of Cas13 proteins is shown in **Table 3.5**.

3.4.8. Cas14

In 2018, a family of CRISPR/Cas systems was discovered; it contained *cas1*, *cas2*, *cas4,* and a new gene *cas14*. Cas14 encodes a small Cas protein (40–70 kDa in

Table 3.4. Variety of Cas12 proteins

No	CRISPR/Cas12a systems	PAM sequence	Reference
1	*Lachnospiraceae bacterium* ND2006 (LbCpf1)	TTTN	[135, 136]
2	*Acidaminococcus* sp. BV3L6 (AsCpf1)	TTTN	[135–137]
3	*Francisella novicida* U112 (FnCpf1)	TTTN	[138]
4	*Candidatus Methanomethylophilus alvus* Mx1201 (CMaCpf1)	TTTV, TTV	[139]
5	*Sneatia amnii* (SaCpf1)	–	[139]
6	*Parcubacteria group bacterium* GW2011 (PgCpf1)	–	[139]
7	*Candidatus Roizmanbacteria bacterium* GW2011 (CRbCpf1)	–	[139]
8	*Candidatus Peregrinbacterium bacterium* GW2011 (CPbCpf1)	–	[139]
9	*Lachnospiracea bacterium* MA2020 (Lb5Cpf1)	TTTV, TTV	[139]
10	*Btyrivibrio* sp. NC3005 (BsCpf1)	TTN, TTTN	[139, 140]
11	*Butyrivibrio fibrisolvens* (BfCpf1)	TTTV	[139]
12	*Prevotella bryantii* B14 (Pb2Cpf1)	TTTV	[139]
13	*Bacteroidetes oral taxon* 274 (BoCpf1)	TTTV, TTV	[139]
14	*Flavobacterium brachiophilum* FL-15 (FbCpf1)	–	[139]
15	*Lachnospiraceae bacterium* MC2017 (Lb4Cpf1)	–	[139]
16	*Moraxella lacunata* (MlCpf1)	TTTV, TTV	[139]
17	*Moraxella bovoculi* AAX08_00205 (Mb2Cpf1)	TTTV, TTN	[139]
18	*Moraxella bovoculi* AAX11_00205 (Mb3Cpf1)	TTTV, TTN	[139]
19	*Thiomicrospira sp.* XS5 (TsCpf1)	TTTV, TTV	[139]
20	*Firmicutes bacterium* ADurb.Bin193 (Adurb193Cas12a)	TTTV	[141]
21	*Archaeon* ADurb.Bin336 (Adurb336Cas12a)	TTTV	[141]
22	*Francisella novicida* (Fn3Cas12a)	YTV	[141]
23	*Prevotella ihumii* (PiCas12a)	KKYV	[141]
24	*Prevotella disiens* (PdCas12a)	TTTV	[141]
25	*Helcococcus kunzii* ATCC 51366 (HkCas12a)	YYV, YYN, TTTN, YTN, TYN, TTYN, TCCN	[140, 141]
26	*Pseudobutyrivibrio ruminis* CF1b (PrCas12a)	TTN, TTTN	[140]
27	*Agathobacter rectalis* strain 2789STDY5834884 (ArCas12a)	TTN, TTTN	[140]
28	*Lachnospira pectinoschiza* strain 2789STDY5834886 (LpCas12a)	TTN, TTTN	[140]
29	*Pseudobutyrivibrio xylanivorans* strain DSM 10317 (PxCas12a)	TTN, TTTN	[140]

Note: N = A, T, G or C; V = A, C, or G; Y = C or T; K = G or T.

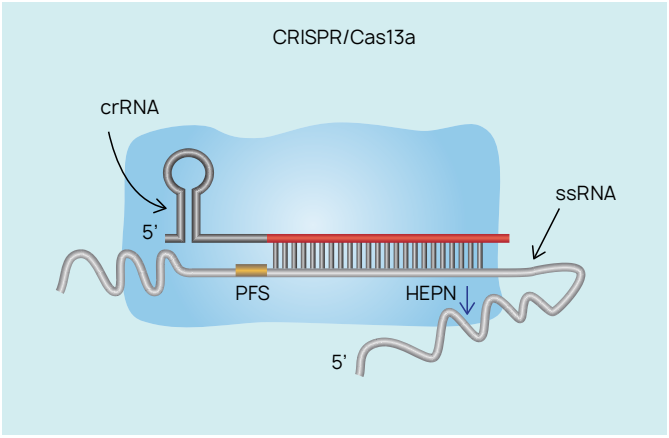


Fig. 3.7. Target RNA and Cas13a-crRNA interaction

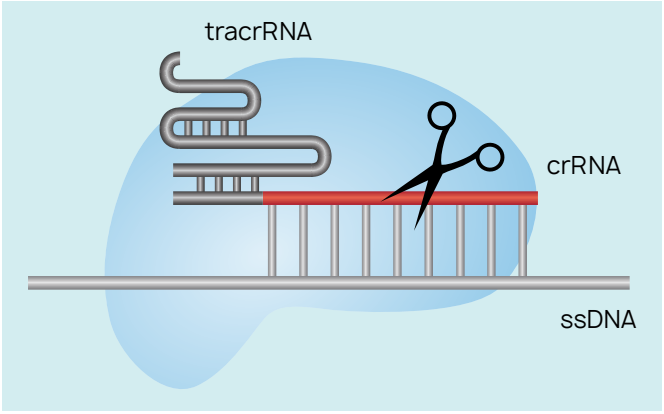


Fig. 3.8. Target ssDNA and Cas14-crRNA+tracrRNA interaction

molecular weight), which is half the size of other Cas proteins found in the so-called class 2 CRISPR/Cas systems [148].

There are 24 variants of the *cas14* gene, which are classified into three subgroups (*cas14a–c*). All these variants share the predicted RuvC nuclease domain typical of a number of CRISPR/Cas enzymes. Unlike other Cas enzymes, Cas14 has not been found in bacterial genomes; it has been found only in the genome of the archaea group. Consequently, it is assumed that Cas14 can be a more primitive version of larger and more complex Cas9 and Cas12 proteins (**Fig. 3.8**).

Cas14 can bind and cleave the target sequence of the single-stranded DNA. Unlike Cas9, Cas14 does not require a PAM sequence. In addition to the specific RNA-guided cleavage, the activated Cas14 can non-specifically cleave a single-stranded DNA [148].

Table 3.5. Diversity of Cas13 proteins

No	CRISPR/Cas13 systems	PFS (protospacer flanking motif)		Reference
		5’	3’	
1	*Leptotrichia shahii* (LshCas13a)	–	H	[40, 144]
2	*Leptotrichia wadei* (LwaCas13a)	–	H	[40, 144]
3	*Listeria seeligeri* (LseCas13a)	–	–	[40]
4	*Lachnospiraceae bacterium* (LbmCas13a)	–	–	[40]
5	*Lachnospiraceae bacterium* (LbnCas13a)	–	–	[40]
6	*Clostridium aminophilum* (CamCas13a)	–	–	[40]
7	*Carnobacterium gallinarum* (CgaCas13a)	–	–	[40]
8	*Carnobacterium gallinarum* (Cga2Cas13a)	–	–	[40]
9	*Paludibacter propionicigenes* (Pprcas13a)	–	–	[40]
10	*Listeria weihenstephanensis* (LweCas13a)	–	–	[40]
11	*Listeriaceae bacterium* (LbfCas13a)	–	–	[40]
12	*Leptotrichia wadei* (Lwa2cas13a)	–	–	[40]
13	*Rhodobacter capsulatus* (RcsCas13a)	–	–	[40]
14	*Rhodobacter capsulatus* (RcrCas13a)	–	–	[40]
15	*Rhodobacter capsulatus* (RcdCas13a)	–	–	[40]
16	*Leptotrichia buccalis* (LbuCas13a)	–	H	[40, 145]
17	*Herbinix hemicellulosilytica* (HheCas13a)	–	–	[40]
18	*Eubacterium rectale* (EreCas13a)	–	–	[40]
19	*Eubacteriaceae bacterium* (EbaCas13a)	–	–	[40]
20	*Blautia* sp. (BmaCas13a)	–	–	[40]
21	*Leptotrichia* sp. (LspCas13a)	–	–	[40]
22	*Bergeyella zoohelcum* (BzoCas13b)	ND	NNNN	[40]
23	*Prevotella intermedia* (PinCas13b)	DD	NNNN	[40]
24	*Prevotella buccae* (PbuCas13b)	ND	NDNN	[40]
25	*Alistipes* sp. (AspCas13b)	DR	NDDN	[40]
26	*Prevotella* sp. (PsmCas13b)	VD	DNNN	[40]
27	*Riemerella anatipestifer* (RanCas13b)	ND	MDDN	[40]
28	*Prevotella aurantiaca* (PauCas13b)	NR	NNNN	[40]
29	*Prevotella saccharolytica* (PsaCas13b)	DN	NNNN	[40]
30	*Prevotella intermedia* (Pin2Cas13b)	DN	NNNN	[40]
31	*Capnocytophaga canimorsus* (CcaCas13b)	ND	NHHA	[40]
32	*Porphyromonas gulae* (PguCas13b)	ND	MHHA	[40]
33	*Prevotella* sp. (PspCas13b)	DD	GBKN	[40]
34	*Flavobacterium branchiophilum* (FbrCas13b)	KR	NNNG	[40]
35	*Porphyromonas gingivalis* (PgiCas13b)	ND	NNNN	[40]
36	*Prevotella intermedia* (Pin3Cas13b)	ND	NNNN	[40]

Table 3.5 to be continued on page 76.

No	CRISPR/Cas13 systems	PFS (protospacer flanking motif)		Reference
		5’	3’	
37	*Fusobacterium necrophorum* (FnsCas13c)	–	–	[40]
38	*Fusobacterium necrophorum* (FndCas13c)	–	–	[40]
39	*Fusobacterium necrophorum* (FnbCas13c)	–	–	[40]
40	*Fusobacterium necrophorum* (FnfCas13c)	–	–	[40]
41	*Fusobacterium perfoetens* (FpeCas13c)	–	–	[40]
42	*Fusobacterium ulcerans* (FulCas13c)	–	–	[40]
43	*Anaerosalibacter* sp. (AspCas13c)	–	–	[40]
44	*Ruminococcus flavefaciens* (RfxCas13d)	–	–	[146]
45	*Eubacterium siraeum* DSM 15702 (EsCas13d)	–	–	[146]
46	*Ruminococcus* sp. N15.MGS-57 (RspCas13d)	–	–	[147]
47	Anaerobic digester metagenome (AdmCas13d)	–	–	[146]
48	*Ruminococcus albus* (RaCas13d)	–	–	[146]
49	*Ruminococcus flavefaciens* FD1 (RffCas13d)	–	–	[146]

Note: D = A, G, T (U); R = A, G; M = A, C; K = G, T (U); B = C, G, T (U); H = A, C, T (U).

3.5. CRISPR/Cas applications

Targeted genome editing with programmable nucleases has taken leading positions among genome modification technologies within a short time. Today, CRISPR/Cas is the most popular system in targeted genome editing [149–151]. CRISPR/Cas nucleases have a number of advantages, namely: high efficiency, multiple editing, low cost, and a short cycle [152, 153].

The CRISPR/Cas systems can have different applications involving gene editing, from cell-based human and animal hereditary disease modeling, functional genome screening, epigenome studies and visualization of cellular processes to applications in the food industry to receive high-quality food products, in agriculture to create new livestock breeds and plant varieties, and in medicine. In addition, the CRISPR/Cas systems can be used in the diagnosis of diseases through the identification of genetic sequences, for example in viruses or oncogenes, for prevention of infectious diseases through gene modification in disease carriers rather than in humans, for example, in malaria, and for the treatment of socially significant diseases, both hereditary and acquired, including cancer, autoimmune diseases, orphan diseases, infectious diseases, and many others.

3.5.1. CRISPR/Cas in the food industry

Fermented foods are a staple in the modern diet, with milk, meat, cucumbers, grains, and cabbage being the most common substrates. Starter cultures play an essential role in transforming these substrates into respective products [154]. Other microbes act as probiotics defined as 'living microorganisms, which, when taken in adequate quantities, are beneficial for the host health' [155]. Probiotics can be incorporated into fermented food products or other food matrices, or they can be available as food supplements. Although fermentation processes including useful microorganisms have been known for centuries, they still encounter such problems as safety and quality of the product, purity and composition of the culture, fermentation failure caused by bacteriophage attack.

Lactic acid bacteria commonly used in starter cultures and probiotics are very frequently found to have CRISPR/Cas systems; their loci are generally found in 62.9% of the studied lactobacilli genomes and 77% of bifidobacteria genomes [156, 157]. The distribution of CRISPR/Cas in these bacteria and the diversity of these systems provide a historical perspective of phago-microbial ecosystems of large-scale fermentation. Besides, CRISPR/Cas can be an excellent tool for fermentation management, including applications for strain typing, phage resistance, plasmid vaccination, genome editing, and anti-microbial activity.

3.5.2. Typing of microorganisms

The adaptive nature of CRISPR/Cas helps receive a dynamic picture of the evolutionary development of a certain microorganism strain. When external threats are detected, new spacers are incorporated into a CRISPR array, in chronological order, at the proximal end of the leader sequence [8, 158, 159]. As the acquisition of new spacers is, first of all, caused by the threats present at some particular time, the position of the spacer in the CRISPR array of the host genome can offer important historical, geographic and environmental information about the particular strain [159–161]. Therefore, CRISPR/Cas becomes an excellent tool not only for strain typing but also for studies in strain divergence and relatedness, microbial ecology and evolution, for epidemiological concepts and understanding of genotypes at the cohort level in complex environmental samples [162, 163].

To identify accurately strains present in starters and probiotics, the food industry needs specific typing techniques. This is essential for maintaining the invariable composition of particular starter culture and for optimizing processes involving isolation and identification of new starter strains to be further used in the industry.

To find a fast and cost-effective way of identification of strain types and variations (typing), a number of techniques were studied and assessed, including pulsed-field gel electrophoresis, PCR-based assay of repeats and 16S rDNA sequencing [164]. Typing by using amplification and sequencing of arrays of repeated CRISPR spacers is a new and efficient tool, which can be added to the existing toolbox. The absence of CRISPR/Cas in the specified strains is the greatest limitation for its application in this field.

Table 3.6. CRISPR/Cas typing of microorganisms

No	Microorganism	Reference
Food industry (starters/probiotics)		
1	*S. thermophilus*	[62]
2	*Lactobacillus casei*	[165]
3	*Lactobacillus paracasei*	[166]
4	*Lactobacillus rhamnosus*	[167]
5	*Enterococcus faecalis*	[168]
6	*Bifidobacterium* genus	[156]
7	*Lactobacillus gasseri*	[169]
Spoliage microorganisms		
8	Spoliage microorganisms	[163]
9	*Lactobacillus buchneri*	[170]
Pathogenic microorganisms		
10	*Salmonella*	[171–182]
11	*Campylobacter jejuni*	[183]
12	*Clostridium difficile*	[184]
13	*Corynebacterium diphtheriae*	[185]
14	*E. coli*	[186]
15	*Legionella pneumophila*	[187]
16	*Staphylococcus aureus*	[188]
17	*Vibrio parahaemolyticus*	[189]
18	*Mycobacterium tuberculosis*	[190, 191]
19	*Yersinia pestis*	[7, 192]

Nevertheless, since CRISPR/Cas is very frequently found in many microorganisms involved in fermentation, the food industry can use the CRISPR/Cas typing to identify species present in the starter cultures and probiotics (**Table 3.6**) [162].

In addition, CRISPR can be useful for tracking pathogenic microorganisms. The speed and easiness of the CRISPR/Cas typing turn it into an ideal method for strain identification in case of contamination problems or disease outbreaks.

The CRISPR/Cas typing can be used independently and in combination with traditional methods used for typing of microorganisms. The application of CRISPR/Cas can provide additional discriminatory power in identifying particular isolates present in starter cultures and probiotics used in the food industry. In addition, the information obtained during CRISPR/Cas typing of food spoilage microorganisms and pathogens can be further used for the development of strategies aimed at the safety and quality of food products.

3.5.3. Constructing microorganism strains resistant to bacteriophages and unwanted plasmid DNAs

Phages are the most abundant biological entities on Earth, outnumbering bacteria [193]. As lytic phages replicate by destroying their bacterial hosts, no wonder that they present a problem for fermentation processes in the food industry. In an attempt

to minimize economic losses caused by each phage infection, the food industry offered several solutions to limit the impact of phages, including changes in the factory design, sanitary measures, ventilation systems, technological protocols, starter media and culture change procedures [194]. Nevertheless, phage infection of starter cultures is the main factor causing slow or failed fermentation in the industry [194]. As phages are often present in raw materials and are robust against different types of treatment, including thermal treatment, high pressure, ionizing radiation and pasteurization, they are virtually impossible to eliminate [195]. Bacteria have evolved to cope with the problem, and many of them have innate mechanisms of protection against bacteriophages [196–201]. CRISPR/Cas is another defense mechanism. Considering the prevalence of phages in fermentation processes, it is not surprising that CRISPR/Cas systems are so common in fermentation microorganisms. CRISPR/Cas helps microorganisms acquire resistance to phages. When a new spacer, which corresponds to the phage it was infected by, is incorporated, the strain is effectively vaccinated against any other encounters with this phage or any other related phages carrying the same protospacer sequence [159].

In addition to their impact on bacteriophages, CRISPR/Cas systems prevent the uptake of plasmids by DNA cleaving [9, 12]. When encountering a plasmid, the microorganism having a CRISPR/Cas system acquires a respective spacer and is subsequently vaccinated against plasmid uptake. The main advantage gained from such vaccination is that it restricts the uptake of undesirable DNA elements, such as pathogenicity islands or antibiotic resistance genes, which are frequently transmitted through plasmids [202–205]. Although vaccination against such elements can develop naturally, the CRISPR/Cas system can also be designed to affect such genetic elements.

Food industry workers are especially concerned about antibiotic resistance prevalent among microorganisms. Monitoring the transfer of genes of these antibiotic-resistance elements in the food chain is seen as a priority in the demonstration of safe application of such microorganisms [206].

In the same manner as phage resistance can be programmed in strains containing native CRISPR/Cas systems, the resistance to plasmids carrying antibiotic resistance genes can be programmed, thus making the modified microorganism resistant to uptake or spread of antibiotic resistance genes [12, 207].

3.5.4. Modification of microorganisms

It is known that CRISPR/Cas9 induces a double-strand break at the specified genomic locus (programmable by a selection of the respective guide RNA). Genome editing takes place when the cell repairs the damage by using endogenous pathways of DNA reparation, such as non-homologous end joining or homologous recombination, which frequently cause alterations in the cleavage site of CRISPR/Cas. However, when dealing with prokaryotes, the approach for using CRISPR/Cas as a tool for genome editing must account for bacterial physiology and DNA homeostasis mechanisms. For example, the double-strand break of bacterial genomes most frequently leads to cell

death because of the lack of robust endogenous DNA repair; therefore, genome editing in bacteria should be driven primarily by recombination events [208]. The interest in using the CRISPR/Cas technique for bacterial genome editing, gene knockout and gene screening in bacterial and archaeal genomes is steadily increasing [209–213].

It should be noted that CRISPR/Cas is an exceptionally well-suited tool for screening and selection of low-frequency desired genotypes [214, 215] as well as for the selection of specific genotypes in fermentation organisms *S. thermophilus* and probiotic *Lactobacillus reuteri* [211, 215]. Besides, the CRISPR/Cas-mediated genome editing technique coupled with recombination and using of linear single-stranded or double-stranded DNA templates has been developed and successfully applied to *E. coli* [216].

Theoretically, CRISPR/Cas can be used to select any number of naturally occurring genotypes for further exploration of their functional abilities and applications, thus making high-throughput genome editing tools useful for selection of economically valuable strains.

3.5.5. Antimicrobial activity

The type I and II CRISPR/Cas systems can be used as programmable antimicrobials; they can easily target undesirable sequences such as antibiotic resistance and virulence genes to destroy pathogenic bacteria or to destroy undesirable plasmids they may sometimes carry [217, 218].

The type II CRISPR/Cas system from *S. pyogenes*, which is characterized by the ability to destroy bacteria communities based on their sequences, was used as a selection tool to introduce mutations [209], providing the first evidence of CRISPR-mediated genome editing in bacteria. The type I systems from *E. coli* and *Salmonella* as well as type II system from *S. thermophilus* were also used for selective destruction of even closely related organisms (up to 99% of the homology) by targeting unique sequences in a complex microbial community [219]. The self-targeting strategy, when the endogenous CRISPR/Cas9 system is used for destroying a bacterial population, makes it possible to eliminate the vast majority of the bacterial population (the destruction efficiency ranges from 2 to 5 folds when using single spacers) [217].

Although the CRISPR self-targeting is a powerful programmable antimicrobial tool, the delivery to the target population remains the main problem in its application for destroying bacteria causing infectious diseases. To achieve clinically significant efficacy, the delivery should be specific and efficient. Some studies showed that the DNA encoding bactericidal proteins different from Cas nucleases can be delivered to bacterial communities by using phage particles as vectors. For example, the M13 phagemid was used for delivery of different toxins or restriction enzymes in *E. coli* [220–222], while the Pf3 phage was also used for delivery of a restriction enzyme and successful treatment of the infection caused by *P. aeruginosa* in mice [223]. With the above strategy, self-guiding CRISPR/Cas systems were successfully delivered to *E. coli* and *S. aureus* by using phage capsids as delivery vectors [224, 225]. In addition, CRISPR/Cas systems with a set of spacers, which are targeted at antibiotic resistance genes, are

successfully used for re-sensitization of cells that initially transfer plasmids with target sequences of antibiotic resistance genes [226].

Thus, the CRISPR/Cas systems are an attractive option for creating programmable and specific antimicrobials. The unique advantage of CRISPR-based antimicrobials over all other strategies (phagotherapy, antimicrobial peptides, antibodies or vaccines) is represented by their ability to kill bacteria carrying strictly specific sequences. This can be useful when it is desirable to destroy only a selected group of bacteria within a particular community, which would be difficult to achieve by using the existing strategies. The application of CRISPR/Cas systems will help solve two main tasks related to the currently available antibiotics. Firstly, it will prevent non-selective elimination of bacteria, which can be useful; secondly, it will reduce the selective burden on the resistance by allowing the off-target population to thrive and occupy the environmental niche. Thus, CRISPR/Cas systems are a novel tool for monitoring the composition of microbial communities rather than a traditional broad-spectrum antibiotic.

3.5.6. HIV therapy

The CRISPR/Cas systems are used to explore target genes and genomic modification processes [227], splicing mechanisms [228], transcription [229] and epigenetic regulation [230]. In addition, CRISPR/Cas systems are used to study and develop therapeutic approaches to the treatment of hereditary diseases [231, 232], infectious diseases, cancers [233] and immunological diseases [234–236].

Targeted genome editing with CRISPR/Cas systems is used as an antiviral therapy when treating infectious diseases. The therapeutic effect is achieved either through alteration of the host genes essential for the life cycle of the virus or through targeting viral genes essential for replication [237]. Today, modification of genes associated with the infection (first of all, with HIV entry into T cells), which is aimed to create $CD4^+$-T cells robust against HIV and to reinfuse the edited cells to patient, is one of the approaches to HIV therapy based on targeted genome editing.

To enter host cells, HIV-1 interacts with the CD4 molecule and CCR5 (C-C chemokine type 5 receptor) or CXCR4 (C-X-C chemokine type 4 receptor) co-receptors. Therefore, HIV tropism is linked to expression patterns of these two co-receptors [238]. There are two types of HIV strains: T-cell-tropic HIV and macrophage-tropic HIV strains. The macrophage-tropic HIV uses the chemokine CCR5 receptor as a co-receptor in the infection of macrophages and primary T cells, and accounts for up to 90% of primary infections. The T-cell-tropic HIV uses CXCR4 as a co-receptor [239–240]. However, it should be noted that there are dual-tropic viruses [238].

The CRISPR/Cas9 systems were used for inducing a site-specific human genomic modification *in vitro* and *in vivo* in mouse models of HIV infection [241–245]. Multiple groups of scientists successfully completed the knockout of the CCR5 receptor in $CD4^+$-T lymphocytes by using CRISPR/Cas9, which was targeted at the open reading frames of the gene encoding CCR5. This approach helped inhibit

HIV-1 infection without significant side effects [244]. Editing of CCR5 both in the population of hematopoietic stem cells and in the population of $CD4^+$-T lymphocytes is a promising strategy for creating HIV-resistant cells and for subsequent re-infusion of edited cells to patients.

Nevertheless, this approach is not efficient for CXCR4-tropic HIV strains. It was found that CRISPR/Cas9 could provide high precision and efficacy in the editing of the CXCR4-encoding gene. The knockout of the HIV CXCR4 co-receptor is accompanied by minor off-target effects and provides resistance to HIV infection caused by CXCR4-tropic HIV strains [242, 246–248]. This approach can be used to create human experimental and therapeutic primary $CD4^+$-T cells to provide an alternative method of treatment of HIV-1 X4 infections. At the same time, the concurrent knockout of both HIV co-receptors — CCR5 and CXCR4 — leads to reduced expression of CCR5 and CXCR4, thus, making modified cells resistant to infection with R5 and X4 tropic viruses even when using dual-tropic viruses [242].

3.5.7. Targeting persistent viral infections

Following the initial infection, many viral pathogens continue to persist in a human body, integrating their genome into a chromosomal DNA or maintaining it episomally within host cells. The viral pathogens that cause persistent infection include HIV, hepatitis viruses, herpesviruses, papillomaviruses, *etc*. In recent years, the CRISPR technology has been successfully used to reduce or eliminate permanent viral infections *in vitro* and in animal models *in vivo*, giving hope that it can be used for the treatment of latent and chronic viral infections [249].

Following acute HIV infection, proviral DNA becomes integrated into host cells, resulting in chronic infection despite antiretroviral therapy. CRISPR/Cas systems were used to combat HIV infection *in vitro* in different cell lines. Researchers were able not only to suppress HIV gene expression in infected T cells and microglial cells but also to eliminate the HIV proviral DNA from multiple other cell lines, including neural progenitor cells, which act as latent reservoirs of HIV infection [250–252].

The CRISPR/Cas systems also showed their efficiency in combating HIV infection *in vivo*. The HIV proviral DNA was eliminated from animals' spleen, lungs, heart, colon, and brain in a humanized model of chronic HIV infection [252]. In addition, researchers used the CRISPR/Cas system to eliminate the HIV proviral DNA from infected human peripheral blood mononuclear cells in a transgenic mouse model [253].

As of today, more than 250 million people are infected with hepatitis B virus (HBV) worldwide, resulting in around 900 thousand deaths each year [254]. The presence of covalently closed circular DNA (cccDNA) of HBV, together with the DNA integrated into the host genome, creates risks of virus reactivation and can lead to the development of hepatocellular carcinoma. Nucleoside/nucleotide analogs per se can hardly eliminate replicative HBV forms composed of cccDNA or integrated HBV DNA. In 2017, the CRISPR/Cas9 system was used to remove a full-length HBV DNA fragment that was chromosomally integrated and episomally located as cccDNA in chronically infected cells. This approach made it possible to completely

eliminate HBV in a stably infected cell line *in vitro*. Thus, it can be assumed that the CRISPR/Cas9 system is a highly promising tool for eradication of HBV chronic infection and complete recovery from HBV [255, 256].

Furthermore, the CRISPR/Cas system has been successfully used *in vitro* in the fight against infections caused by herpesviruses. Guide RNAs used simultaneously led to a significant reduction in replication of herpes simplex virus 1 in cells [257, 258]. The CRISPR/Cas system can eliminate up to 95% of the Epstein–Barr virus and cytomegalovirus DNA within 11 days; then, the organism develops mutant viral forms resistant to CRISPR/Cas [257, 258]. The efficiency of CRISPR/Cas systems was also demonstrated by eliminating other viral pathogens *in vitro*, such as John Cunningham virus and human papillomavirus 16 and 18 [259, 260].

3.5.8. Development of therapeutic approaches to monogenic diseases

Monogenic diseases are caused by a defect in a single gene and are inherited according to traditional Mendelian patterns [261]. These disorders affect millions of people, and it has been estimated that more than 10 thousand human diseases fall under this category. Monogenic diseases are mainly classified as dominant, recessive, and X-linked [261]. Treatment of most of these diseases is still limited to the management of symptoms without addressing the underlying genetic defect. The advent of genome editing tools such as the CRISPR/Cas system opens avenues for the development of therapeutic strategies for the treatment of monogenic diseases (**Table 3.7**) [262].

3.5.9. Development of therapeutic approaches to cancer treatment

Cancer is responsible for millions of human deaths worldwide, and researchers have been searching for efficient methods of treatment for years. New drugs, chemotherapy, and radiological therapy have been available for the treatment of patients for years; however, all the offered approaches have side effects. The CRISPR/Cas revolutionary tool for genome editing opened up new vistas for developing therapeutics for cancer. CRISPR/Cas has enormous potential for advancement in gene and cell therapy for malignancies (**Table 3.8**).

3.6. Clinical trials

Currently, a total of 30 clinical trials have been launched to evaluate candidate therapeutic products based on CRISPR/Cas nucleases. The CRISPR/Cas efficiency is evaluated for hematological and solid malignancies such as acute lymphoblastic leukemia, non-Hodgkin lymphoma, lymphoblastic lymphoma, esophageal cancer, invasive bladder cancer, hormone-resistant prostate cancer, and non-small cell lung cancer as well as for hereditary diseases such as sickle cell disease, thalassemia, Kabuki syndrome, Rubinstein-Taybi syndrome. Finally, as of June 2020, there are two clinical trials evaluating CRISPR/Cas designed to combat infectious diseases, such as HIV and human papillomavirus (**Table 3.9**).

Table 3.7. CRISPR/Cas-mediated therapeutic approaches to human monogenic diseases

No	Target & Strategy	Outcome	Reference
Cystic fibrosis			
1	Strategy: CRISPR/Cas9, guide RNA specific to the *CFTR* gene sequence Method: homologous recombination Delivery: adeno-associated viral vector	Successful editing of Delta F508 *CFTR* mutation in easily accessible airway basal stem cells obtained from patients with cystic fibrosis. The editing resulted in 30–50% allelic correction in stem cells and bronchial epithelial cells in 10 patients (the *CFTR* function was restored to 20–50% against the control samples without cystic fibrosis in differentiated epithelium). The modified cells retained their differentiation capability, as shown on the animal model *in vivo* (the pig)	[263]
2	Strategy: CRISPR/Cas9, guide RNA specific to the *CFTR* gene sequence. Method: homologous recombination Delivery: plasmid DNA	CRISPR/Cas9-mediated homologous recombination successfully corrects the mutant F508del allele in intestinal stem cells isolated from two patients with cystic fibrosis. The functionality of the corrected allele is demonstrated in the organoid system	[264]
3	Strategy: CRISPR/Cas9, guide RNA specific to the *CFTR* gene sequence. Method: homologous recombination Delivery: plasmid DNA	The CRISPR system is used to correct deletion of F508 in the *CFTR* gene in induced pluripotent stem cells (iPSCs) obtained from patients with cystic fibrosis. It was observed that the corrected iPSCs had normal CFTR expression and function when they differentiated into mature airway epithelial cells.	[265]
Sickle cell disease			
4	Strategy: CRISPR/Cas9, guide RNA specific to the *HBB* gene sequence. Method: homologous recombination Delivery: ribonucleoprotein complex	It has been found that HiFi Cas9 provides efficient correction (homologous recombination) of the Glu6Val mutation in the *HBB* gene, which causes sickle cell disease, in $CD34^+$ cells obtained from patients with sickle cell disease	[266]
5	Strategy: CRISPR/Cas9 hGemCas9, guide RNA specific to the *HBB* gene sequence. Method: homologous recombination Delivery: mRNA	The modified version of Cas9 (hGemCas9) with reduced nuclease activity in the G1 phase of the cell cycle, and synchronization of mobilized peripheral blood stem cells in S/G2 phases resulted in a 4-fold increase in the rate of homologous recombination *in vitro* and *in vivo*	[267]

Table 3.7 to be continued on page 85.

No	Target & Strategy	Outcome	Reference
6	Strategy: CRISPR/Cas9, guide RNA specific to the *HBB* gene sequence. Method: homologous recombination Delivery: information is not available	The novel editing technology in human pluripotent cells by using the Cas9 protein combined with chemically modified guide RNAs and recombinant AAV6 vectors for delivery of donor templates for homologous recombination can be used for integration of a 2.2-thousand bp DNA expression cassette at frequencies up to 94% at the *HBB* locus. The technology was used successfully to correct 63% of the iPSCs obtained from patients with sickle cell disease	[268]
Thalassemia			
7	Strategy: CRISPR/Cas9, guide RNA specific to the HbE mutation sequence Method: homologous recombination Delivery: plasmid DNA	The mutation in the β-globin gene (HBB) in iPSCs obtained from patients with HbE/β-thalassemia was corrected by using the CRISPR/Cas9 system. Modified clones were differentiated into erythroid cells that contained mature *HBB* gene and expressed functional HBB protein	[269]
8	Strategy: CRISPR/Cas9, guide RNA specific to $HBB^{IVS\text{-}110(G>A)}$ mutation sequence Method: non-homologous end joining Delivery: ribonucleoprotein complex	CRISPR/Cas9 corrects, at 95% efficiency, the $HBB^{IVS\text{-}110(G>A)}$ mutation in CD34^{+}-$HBB^{IVS\text{-}110(G>A)}$ homozygous erythroblasts obtained from patients	[270]
9	Strategy: CRISPR/Cas9, guide RNA specific to the *HBB CD41/42(−CTTT)* mutation sequence Method: homologous recombination Delivery: lentiviral vector	CRISPR/Cas9-ssODNs successfully correct the CD41/42 (-CTTT) mutation of the the β-globin gene (HBB) in iPSCs obtained from patients with β-thalassemia. Modified clones retain full pluripotency and have normal karyotypes. Erythroblasts differentiated from modified iPSCs express *HBB*	[271]
Huntington's Disease (HD)			
10	Strategy: CRISPR/Cas9, guide RNA specific to mutant huntingtin (mHTT) sequence Method: knockdown Delivery: adeno-associated viral vector	The strategy for allele-specific editing of mHTT sequence based on CRISPR/Cas9 technology takes advantage of highly prevalent SNPs at the *HTT* locus for guiding mutant allele-specific cleavage and shows its effectiveness in reducing the expression of mutant proteins in human HD fibroblasts *in vitro* and mice brain *in vivo*	[272]

Table 3.7 to be continued on page 86.

No	Target & Strategy	Outcome	Reference
11	Strategy: CRISPR/Cas9, guide RNA specific to mutant huntingtin (mHTT) sequence Method: knockdown Delivery: adeno-associated viral vector	The reduction of mHTT expression in striatal neuronal cells in adult mice (specialized HD model) did not affect viability, though alleviated motor deficit. Studies showed that CRISPR/Cas9-medicated allele-specific editing could be used to eliminate efficiently neuronal cell toxicity caused by an extended polyglutamine (polyQ) tract, in the human adult brain	[273]
12	Strategy: CRISPR/Cas9 KamiCas9 (a self-inactivating editing system for achieving transient expression of the Cas9 protein and high editing efficiency), guide RNA (sgHTT1) targeted at the region close to the *HTT* translation start site to cause permanent *HTT* disruption Method: knockout Delivery: lentiviral vector	The KamiCas9 system demonstrated high efficiency in genome editing of neuronal and glial cells of the mouse brain and in iPSC cultures obtained from HD patients. The molecular analysis demonstrated an improved safety profile of KamiCas9, which is essential in the context of CNS applications and, in particular, slowly progressive neurodegenerative diseases such as Huntington's disease	[274]
13	Strategy: CRISPR/Cas9 from *Staphylococcus aureus*, guide RNA specific to mutant huntingtin (mHTT) sequence Method: knockout Delivery: adeno-associated viral vector	The *Staphylococcus aureus* Cas9 nuclease packaged with a single guide RNA in the adeno-associated viral vector can be used to disrupt the expression of the mutant *HTT* gene in the mouse model of Huntington's disease following its *in vivo* delivery to the striatum. It was found that CRISPR-Cas9-mediated disruption of the mutant *HTT* gene resulted in a 50% decrease in neuronal inclusions and significantly improved lifespan and certain motor deficits	[275]
Duchenne muscular dystrophy (DMD)			
14	Strategy: CRISPR/Cas9 from *Staphylococcus aureus*, guide RNA specific to intron 22 and 23 sequence (to remove the exon 23 from the *Dmd* gene in the MDX mouse model of DMD) Method: non-homologous end joining Delivery: adeno-associated viral vector	AAV-CRISPR/Cas9 successfully removes the mutated exon 23 from the dystrophin gene. The deletion of the exon 23 restores expression of a modified dystrophin gene, partial recovery of functional dystrophin protein in skeletal myofibers and cardiac muscle, improvement of muscle biochemistry, and significant enhancement of muscle force in model animals	[276]

Table 3.7 to be continued on page 87.

No	Target & Strategy	Outcome	Reference
15	Strategy: CRISPR/Cas9 from *Staphylococcus aureus*, guide RNA specific to intron 22 and 23 sequence (to remove the exon 23 from the *Dmd* gene in the MDX mouse model of DMD) Method: non-homologous end joining Delivery: adeno-associated viral vector	AAV-CRISPR restores dystrophin expression in the mouse model of DMD for one year. AAV-CRISPR did not cause any signs of toxicity in the mouse model during one year	[277]
16	Strategy: CRISPR/Cas9 from *Streptococcus pyogenes*, guide RNA specific to sequences of splice acceptor or donor sites of exons 43 and 45 of the *Dmd* gene in the mouse model of DMD Method: non-homologous end joining Delivery: adeno-associated viral vector	The simple and efficient strategy is designed for correction of exon 44 deletion mutations by CRISPR/Cas9 gene editing in cardiomyocytes obtained from patients' iPSCs *in vitro* and *in vivo*	[278]
17	Strategy: CRISPR/Cas9 from *Streptococcus pyogenes* and *Staphylococcus aureus*, guide RNA specific to the *Dmd* gene sequence Method: non-homologous end joining, homologous recombination Delivery: adeno-associated viral vector	AAV-CRISPR/Cas9 edits the *Dmd* gene in the mdx4cv mouse model. Treated muscles demonstrate robust dystrophin expression following both local and systemic delivery, resulting in significant morphometric and pathophysiological improvement of the dystrophic phenotype. AAV-CRISPR/Cas9 induces *Dmd* gene correction *in vivo* through homologous recombination	[279]
18	Strategy: CRISPR/Cas9 from *Streptococcus pyogenes*, guide RNA specific to the *Dmd* gene sequence (the exon 51 splice acceptor site) Method: non-homologous end joining Delivery: adeno-associated viral vector	AAV-CRISPR/Cas9 restores dystrophin expression in dogs 6 weeks after intramuscular delivery or 8 weeks after systemic delivery. After the systemic delivery in skeletal muscles, dystrophin was restored to levels ranging from 3 to 90% of the normal level, depending on the muscle type. In the cardiac muscle, the dystrophin levels in the dogs receiving the highest dose reached 92% of the normal level	[280]

Table 3.7 to be continued on page 88.

No	Target & Strategy	Outcome	Reference
19	Strategy: CRISPR/Cpf1 from *Lachnospiraceae bacterium* and *Acidaminococcus*, guide RNA specific to the *Dmd* gene sequence Method: non-homologous end joining, homologous recombination Delivery: adeno-associated viral vector	Cpf1 efficiently corrects mutations causing Duchenne muscular dystrophy, *in vitro* in human cells and *in vivo* in the mouse model	[281]
Hemophilia			
20	Strategy: CRISPR/Cas9, guide RNA specific to the sequence of the *ROSA26* gene encoding clotting factor *FIX* Method: homologous recombination Delivery: adenoviral vector	The adenoviral vector delivery of CRISPR/Cas9 components and templates for homologous recombination provides correct restoration of the FIX-encoding gene, thus contributing to long-term improvement in FIX activity and phenotypic correction in the mouse model of juvenile hemophilia	[282]
21	Strategy: CRISPR/Cas9, guide RNA specific to the sequence of the gene encoding clotting factor FVIII (exon 14) Method: homologous recombination Delivery: plasmid DNA	CRISPR/Cas9-ssODN efficiently corrects the *FVIII*-encoding gene in iPSCs obtained from patients with hemophilia (HA-iPSCs). The *FVIII* expression and activity were restored *in vitro* and *in vivo* in endothelial progenitor cells obtained from modified HA-iPSC	[283]
22	Strategy: CRISPR/Cas9, guide RNA specific to the sequence of the integration locus of the adeno-associated virus (AAVS1) Method: homologous recombination Delivery: plasmid DNA	The encoding sequence of human F9 was integrated into the AAVS1 locus in iPSCs obtained from the hemophilia B patient by using the CRISPR/Cas9 system. The hepatocytes obtained during differentiation of modified iPSCs expressed steadily F9, even after implantation to animals *in vivo*	[284]
23	Strategy: CRISPR/Cas9 *Staphylococcus aureus*, guide RNA specific to the intron sequence 13 at the locus of liver-specific albumin (Alb) Method: homologous recombination Delivery: adeno-associated viral vector	The genome editing technology used dual AAV vectors encoding *Staphylococcus aureus* Cas9, guide RNA (SaCas9-gRNA) and codon-optimized gene encoding human FVIII with human B-domain deletion (BDD-F8). BDD-F8, in a site-specific manner, is integrated into the locus of the liver-specific albumin (Alb), thus causing production of FVIII in the liver. In the mouse FVIII-gene knocked out model (F8KO, hemophilia A model), editing caused increased levels of FVIII protein, which remained active in the liver within 7 months, without any noticeable toxicity for the liver	[285]

Table 3.7 to be continued on page 89.

No	Target & Strategy	Outcome	Reference
Diabetes			
24	Strategy: dCas9/sgFabp4 Method: CRISPRi Delivery: adipocyte-specific peptide CKGGRAKDC and polyarginine (9 residues) — ATS-9R	The targeted delivery of the anti-Fabp4 CRISPRi system to white adipocytes by using ATS-9R resulted in the effective Fabp4 knockout, thus causing a reduction in the body weight, reduction in inflammation, and restoration of hepatic function in mice with obesity. The application of anti-Fabp4 CRISPRi system ameliorated obesity and type 2 diabetes caused by obesity through suppression of the Fabp4 expression	[286]
25	Strategy: CRISPR/Cas9+ssODN Method: homologous recombination Delivery: information is not available	Pancreatic β-cells obtained from iPSCs and CRISPR/Cas9-modified for correction of the pathogenic variant causing diabetes, in *Wolfram syndrome* type *1* (WFS1) secreted insulin *in vitro* in response to glucose and stopped the development of streptozotocin-induced diabetes after grafting to mice	[287]
26	Strategy: CRISPR/Cas9/sgDPP-4 Method: knockout Delivery: nanoparticles	The delivery of CRISPR/Cas9/sgDPP-4 ribonucleoprotein complexes specific to dipeptidyl peptidase-4 gene modulates the function of glucagon-like peptide 1. Complexes injected with nanoparticles to model mice with insulin-dependent type 2 diabetes reduced the blood glucose level, normalized the response to insulin and reduced damage in the liver and kidneys	[288]
27	Strategy: CRISPR/Cas9, CRISPR/Cas9, guide RNA specific to the sequences encoding peptidyl-glycine alpha-amidating monooxygenase (PAM), an insulin destroying enzyme (IDE) and insulin (INS) Delivery: lentiviral vector	Successful creation of knockout lines of pancreatic β-cells (EndoC-βH1) demonstrate the possibility of CRISPR/Cas9-based genome editing in the above types of cells and offers additional opportunities for using of CRISPR interference (CRISPRi), CRISPR activation (CRISPRa), CRISPR genome screening systems, epigome modification systems and CRISPR base editing systems	[289]
Cardiovascular diseases			
28	Strategy: CRISPR/Cas9, guide RNA specific to sequences encoding TNNI3K (TNNI3 interacting kinase) Method: knockout Delivery: plasmid DNA	Excessive and/or normal TNNI3K expression is associated with susceptibility to dilated cardiomyopathy. It is assumed that TNNI3K knockout can prevent ventricular dilation, which will be clinically beneficial for patients with refractory disease	[290]

Table 3.7 to be continued on page 90.

No	Target & Strategy	Outcome	Reference
29	Strategy: CRISPR/Cas9, guide RNA specific to sequences encoding the low-density lipoprotein receptor (LDLR) Method: homologous recombination Delivery: adeno-associated viral vector	AAV-CRISPR/Cas9 provides *in vivo* correction of the *Ldlr* gene and partially restores the LDLR expression, thus reducing symptoms of atherosclerosis in model animals. AAV-CRISPR/Cas9 can offer an efficient therapeutic approach to the treatment of patients with familial hypercholesterolemia	[291]
30	Strategy: CRISPR/Cas9, guide RNA specific to the sequence of adeno-associated virus (AAVS1) integration locus Method: homologous recombination Delivery: plasmid DNA	The CRISPR/Cas9 system helps restore wild-type *LDLR* (normal phenotype) expression in the iPSC model. The obtained model cells can be used for studying the regulation of cholesterol metabolism. It was found that *LDLR* plays role in the late stage of the hepatitis C virus (HCV) life cycle — in assembly or secretion of a viral progeny. $LDLR^{+}$-iPSC can be used as a platform for screening of drugs administered for treatment of dislipidemy and HCV infection	[292]
31	Strategy: CRISPR/Cas9, guide RNA specific to the sequence encoding proprotein convertase subtilisin/kexin type 9 (Pcsk9) Method: knockout Delivery: adenoviral vector	The study has proved efficacy and safety of the CRISPR/Cas9 system targeted at human *PCSK9* gene in human hepatocytes *in vivo* (chimeric mice with a humanized liver)	[293]

Table 3.8. CRISPR/Cas-mediated therapeutic approaches to human cancers

No	Target & Strategy	Outcome	Reference
1	Strategy: CRISPR/Cas9, respective guide RNAs Method: TCR, B2M and PD-1 molecules were removed simultaneously to increase anti-tumor activity. Other genes, such as *CTLA-4*, *LAG-3*, *TIM-3* and *Fas*, were also destroyed together with TCR and B2M Delivery: different techniques	Creating CAR-T cells – T cells with chimeric antigen receptor, having high anti-tumor activity, including universal CAR-T cells — allogeneic T cells with removed endogenous TCR and HLA	[294–304]
2	Strategy: CRISPR/Cas9, respective guide RNAs Method: homologous recombination Delivery: different techniques	CAR-T cells, in which the CAR or TCR-cassette is inserted in the endogenous locus of the *TCR* gene to alleviate the graft-versus-*host response*, thus making impossible any accidental integration of the cassettes and providing smooth expression of CAR (chimeric antigen receptor)	[305–309]

Table 3.9. Clinical trials of candidate therapeutic products based on CRISPR/Cas nucleases

No	Title	*ClinicalTrials.gov* identifier	Application	Start/approval date	Location
Oncology					
1	A Phase I Clinical Trial of PD-1 Knockout Engineered T Cells Treating Patients With Advanced Non-small Cell Lung Cancer	NCT02793856	Metastatic non-small cell lung cancer	August 2016	China
2	A Safety and Efficacy Study of Transcription Activator-like Effector Nucleases and Clustered Regularly Interspaced Short Palindromic Repeat/ Cas9 in the Treatment of HPV-related Cervical Intraepithelial Neoplasia	NCT03057912	Cervical intraepithelial neoplasia	July 2017	China
3	Safety and Activity of Programmed Cell Death-1 Knockout Engineered T Cells in Patients With Previously Treated Advanced Oesophageal Squamous Cell Carcinoma: An Open-label, Single-arm Phase 1 Study	NCT03081715	Esophageal cancer	March 2017	China
4	A Dose-escalation Phase I Trial of PD-1 Knockout Engineered T Cells for the Treatment of Muscle-invasive Bladder Cancer	NCT02863913	Muscle-invasive bladder cancer	September 2016	China
5	A Dose-escalation Phase I Trial of PD-1 Knockout Engineered T Cells for the Treatment of Castration Resistant Prostate Cancer	NCT02867345	Prostate cancer	November 2016	China
6	A Dose-escalation Phase I Trial of PD-1 Knockout Engineered T Cells for the Treatment of Metastatic Renal Cell Carcinoma	NCT02867332	Metastatic renal cell carcinoma	November 2016	China

Table 3.9 to be continued on page 92.

No	Title	*ClinicalTrials.gov* identifier	Application	Start/approval date	Location
7	Safety and Feasibility Study of Allotransplantation of CRISPR/ Cas9 CCR5 Gene Modified CD34+ Haematopoietic Stem/Progenitor Cells in HIV-infected Subjects With Hematological Malignancies	NCT03164135	Hematological malignances	May 2017	China
8	Phase I/II Study to Determine the Safety, Tolerability, Biological Activity and Efficacy of Universal CRISPR-Cas9 Gene-Editing CAR-T Cells Targeting CD19(UCART019) in Patients With Relapsed or Refractory CD19+ Leukaemia and Lymphoma	NCT03166878	Leukemia and lymphoma	May 2017	China
9	Phase I Study to Evaluate Treatment of CRISPR-Cas9 Mediated PD-1 and TCR Gene-knocked Out Chimeric Antigen Receptor (CAR) T Cells in Patients With Mesothelin Positive Multiple Solid Tumors	NCT03545815	Solid tumors	March 2018	China
10	Phase I Study of CRISPR-Cas9 Mediated PD-1 Gene-knocked Out Mesothelin-directed CAR-T Cells With the Conditioning Regimen of Paclitaxel and Cyclophosphamide in Mesothelin Positive Multiple Solid Tumors	NCT03747965	Solid tumors	November 2018	China

Table 3.9 to be continued on page 93.

No	Title	*ClinicalTrials.gov* identifier	Application	Start/approval date	Location
11	Evaluation of the safety of CRISPR/Cas9-mediated HPK1 gene-knocked out CD19-targeted CAR-T cells (T cells with the chimeric antigen receptor) used for treatment of recurrent or refractory hematopoietic malignancies (Phase I)	Information is not available	Hematopoietic CD19-positive malignancies	August 2019	China
12	A Phase I/II Trial of PD-1 Knockout EBV-CTLs for Advanced Stage EBV Associated Malignancies	NCT03044743	Stage IV gastric carcinoma, stage IV nasopharyngeal carcinoma, stage IV T-cell lymphoma, stage IV adult Hodgkin lymphoma, stage IV diffuse large B-cell lymphoma	March 2017	China
13	A Phase 1 Dose Escalation and Cohort Expansion Study of the Safety and Efficacy of Anti-BCMA Allogeneic CRISPR-Cas9-Engineered T Cells (CTX120) in Subjects With Relapsed or Refractory Multiple Myeloma	NCT04244656	Multiple myeloma	January 2020	Multi-center: USA, Australia, Spain
14	A Phase 1 Dose Escalation and Cohort Expansion Study of the Safety and Efficacy of Allogeneic CRISPR-Cas9-Engineered T Cells (CTX130) in Subjects With Advanced, Relapsed or Refractory Renal Cell Carcinoma With Clear Cell Differentiation	NCT04438083	Renal cell carcinoma	June 2020	Australia

Table 3.9 to be continued on page 94.

No	Title	*ClinicalTrials.gov* identifier	Application	Start/approval date	Location
15	A Phase 1 Dose Escalation and Cohort Expansion Study of the Safety and Efficacy of Allogeneic CRISPR-Cas9-Engineered T Cells (CTX110) in Subjects With Relapsed or Refractory B-Cell Malignancies (CARBON)	NCT04035434	B-cell malignancy, non-Hodgkin lymphoma, B-cell lymphoma	July 2019	Multi-center: USA, Australia, Germany
16	Phase I/II Study to Evaluate Treatment of Relapsed or Refractory Leukemia and Lymphoma With Universal CRISPR-Cas9 Gene-Editing CAR-T Cells Targeting CD19 and CD20 or CD22	NCT03398967	B-cell leukemia, B-cell lymphoma	January 2018	China
17	Cell Therapy for High Risk T-Cell Malignancies Using CD7-specific CAR Expressed On Autologous T Cells (CRIMSON)	NCT03690011	T-cell acute lymphoblastic leukemia, T-cell acute lymphoblastic lymphoma, T-non-Hodgkin lymphoma	October 2018	USA
18	Safety and Effect Assessment of TACE in Combination With Autologous PD-1 Knockout Engineered T Cells by Percutaneous Infusion in the Patients With Advanced Hepatocellular Carcinoma	NCT04417764	Hepatocellular carcinoma	June 2020	China
19	Evaluation of the safety and efficacy of CRISPR/Cas9-engineered, cytokine-induced SH2 protein gene-knocked out tumor infiltrating lymphocytes for treatment of patients with gastrointestinal cancer (Phase I/II)	Information is not available	Gastrointestinal cancer	May 2020	USA
Monogenic disorders					

Table 3.9 to be continued on page 95.

No	Title	*ClinicalTrials.gov* identifier	Application	Start/approval date	Location
20	A Phase 1/2 Study of the Safety and Efficacy of a Single Dose of Autologous CRISPR-Cas9 Modified CD34+ Human Hematopoietic Stem and Progenitor Cells (hHSPCs) in Subjects With Transfusion-Dependent β-Thalassemia	NCT03655678	β-Thalassemia	September 2018	Multi-center: UK, Germany
21	A Phase 1/2 Study to Evaluate the Safety and Efficacy of a Single Dose of Autologous CRISPR-Cas9 Modified CD34+ Human Hematopoietic Stem and Progenitor Cells (CTX001) in Subjects With Severe Sickle Cell Disease	NCT03745287	Sickle cell disease	October 2018	Multi-center: Germany, Italy, UK, France, Belgium
22	Open-Label, Single Ascending Dose Study to Evaluate the Safety, Tolerability, and Efficacy of EDIT-101 in Adult and Pediatric Participants With Leber Congenital Amaurosis Type 10 (LCA10), With Centrosomal Protein 290 (CEP290)-Related Retinal Degeneration Caused by a Compound Heterozygous or Homozygous Mutation Involving c.2991+1655A>G in Intron 26 (IVS26) of the CEP290 Gene ("LCA10-IVS26")	NCT03872479	Blindness, Leber congenital amaurosis type 10, visual impairment, eye diseases, hereditary eye disorders, retinal diseases, retinal degeneration	September 2019	USA

Table 3.9 to be continued on page 96.

No	Title	*ClinicalTrials.gov* identifier	Application	Start/approval date	Location
23	Diagnosis of the Rubinstein–Taybi syndrome: identification of acetylation profiles as epigenetic markers for assessing causality of CREBBP (GENEPI) variants, including the transcriptome analysis with RNA-Seq isogenic clones of stem cells obtained through CRISPR/Cas9-mediated correction of CREBBP mutations in patients with Rubinstein–Taybi syndrome	Information is not available	Rubinstein–Taybi syndrome	October 2019	France
24	A Long-term Follow-up Study of Subjects With β-thalassemia or Sickle Cell Disease Treated With Autologous CRISPR-Cas9 Modified Hematopoietic Stem Cells (CTX001)	NCT04208529	β-Thalassaemia, sickle cell disease, hematologic diseases, hemoglobinopathies, inborn genetic diseases	December 2019	Multi-center: USA, Germany, Italy
25	A Safety and Efficacy Study of a Single Center, Open-label, Single Arm About the Gene Correction of HBB in Patient-specific iHSCs Using CRISPR/Cas9 That Intervent Subjests With β-thalassemia Mutations	NCT03728322	β-Thalassaemia	January 2019	Not provided
26	Exploiting Epigenome Editing in Kabuki Syndrome: a New Route Towards Gene Therapy for Rare Genetic Disorders	NCT03855631	Kabuki syndrome	February 2019	France

There is no doubt that the number of clinical trials of CRISPR/Cas-modified therapeutic products is going to increase from year to year, as numerous candidate therapeutics are being evaluated through preclinical studies for such conditions as infectious diseases, orphan diseases, cancers.

3.7. CRISPR/Cas delivery systems

The efficient delivery of CRISPR/Cas genome editing elements to target cells is of paramount importance for using CRISPR/Cas tools in therapy [310]. Generally, three strategies of delivery of CRISPR/Cas elements are used — *in vitro*, *ex vivo* and *in vivo*, as well as other methods such as physical methods, viral and non-viral vector delivery, *etc*. Physical methods of delivery imply short-term disruption of the membrane target cell and include electroporation, sonoporation, nano-injection, microinjection, and hydrodynamic injection [311]. Viral vectors are the earliest molecular tool for transfer of genes to human cells; they transfer nucleic acids encoding CRISPR/Cas components to target cells in the envelope of a virus, for example, an adenovirus, adeno-associated virus, retrovirus, lentivirus, Epstein–Barr virus, herpes simplex virus and bacteriophages [312, 313]. In addition, the recently reported alternative (non-viral) methods of CRISPR/Cas delivery, for example, by using lipid nanoparticles, polymer nanoparticles and hydrogel nanoparticles, hybrid gold, graphene oxide, metal-organic framework, black phosphorus nanomaterials, *etc.* [314].

CRISPR/Cas elements can be delivered to a live cell as a set of plasmid DNAs encoding the Cas protein and guide RNA or as a combination of the Cas-protein-encoding mRNA and the guide RNA. The third option suggests delivery of the pre-assembled ribonucleoprotein complex (Cas protein and the guide RNA) into the cell (**Table 3.10**). The CRISPR/Cas delivery in the form

Table 3.10. CRISPR/Cas delivery systems

Method of delivery	Forms of delivered CRISPR/Cas elements		
	DNA	mRNA	protein
Electroporation	+	+	+
Viral vectors	±	±	–
Lipofection	+	+	+
Lipid nanoparticles	–	+	+
Polymer nanoparticles	–	–	+
Hydrogel nanoparticles	–	–	+
Gold nanoparticles	–	–	+
Graphene oxide	–	–	+
Metal-organic compounds	–	–	+
Nanolayered black phosphorus	–	–	+
Cell penetrating peptides	–	–	+
DNA nanostructures	–	–	+

of a ribonucleoprotein complex has several advantages, including high editing efficiency; low non-specific activity; editing starts immediately after the delivery to the cell; fast screening of efficiency of guide RNAs *in vitro*; reduced immunogenicity due to the short-term presence of CRISPR/Cas elements in the target cell. Therefore, ribonucleoprotein complexes offer promising opportunities in CRISPR/Cas-based genome editing.

3.8. Diagnosis of infectious diseases

Solutions of epidemiological problems addressing the understanding of infectious disease outbreaks, detection and identification of the pathogen, and detection of specific bacterial genes require design and implementation of advanced technologies of molecular epidemiology. One of such technologies involves using CRISPR/Cas genetic editing elements. This technology is successfully used for developing treatment strategies for some diseases, despite some difficulties associated with unforeseen mutations. Thorough studies of the CRISPR/Cas system found that it could be used for fine diagnostic procedures aimed to identify the pathogen(s) of infection in humans and to genotype them.

3.8.1. CRISPR/Cas9-based diagnosis

CRISPR/Cas9 was used by several researchers developing diagnostic kits for identification of infectious diseases. For example, to detect Zika virus, K. Pardee and collaborators first amplified the viral RNA using an isothermal amplification — nucleic acid sequence based amplification (NASBA) combined with the CRISPR/Cas9 system. By using the technique, scientists could accurately differentiate between closely related viral strains by the presence/absence of the strain-specific PAM [315].

In 2016, the technique based on optical mapping was offered for direct identification of antibiotic resistance genes. In this technique, DNA sequences of single plasmids, which carry antibiotic resistance genes and are present in bacterial isolates in nanofluidic channels, were optically mapped. The technique makes it possible to identify antibiotic resistance genes by using CRISPR/Cas9 and guide RNAs specific to antibiotic resistance genes, such as *blaCTX-M-1* group, *blaCTX-M-9* group, *blaNDM,* and *blaKPC*. During the assay, the CRISPR/Cas9 ribonucleoprotein complex linearizes circular plasmids in the antibiotic resistance gene region, and the resulting linear DNA molecules are identified by using optical DNA mapping. In future, the offered assay and technique can be applied to low concentration samples to identify antibiotic resistance genes [316].

Later, CRISPR/Cas9 was combined with FISH (fluorescence *in situ* hybridization) to detect methicillin-resistant *Staphylococcus aureus* (MRSA) strains. The method employs the dCas9 system, in which the ribonucleoprotein complex coupled with a SYBR fluorescent probe recognizes the *mecA* gene of *S. aureus*. The method makes it possible to detect MRSA at a concentration of 10 CFU/ml and to distinguish between *S. aureus* isolates with and without the *mecA* gene [317].

In addition to the above-mentioned technologies, a new technology was reported in 2019. It combined next-generation sequencing (NGS) and CRISPR/Cas9 capabilities. The novel technology was called Finding Low Abundance Sequences by Hybridization (FLASH) and is used for targeted enrichment during NGS. The FLASH method uses a set of guide RNAs that cut the intended sequences into fragments suitable for further sequencing by using the Illumina platform. The input genomic DNA or cDNA is first blocked by phosphatase treatment and then is cleaved with Cas9 ribonucleoprotein complexes programmed by the set of guide RNAs. The resulting cleavage products are ligated to universal adapters for sequencing. During the subsequent amplification, the target sequences are enriched over the background and get ready for binding to the sequencing flow cell. This method goes beyond other CRISPR-based diagnostic tools by providing high levels of multiplexing and making it possible to analyze concurrently up to 1,000 targets, being reinforced by the precision and sequence identity inherent in a traditional NGS readout. FLASH-NGS was successfully used in studies and diagnosis of antibiotic-resistant infections to evaluate the burden of antimicrobial resistance genes in pneumonia-causing gram-positive bacteria and drug resistance in the malaria parasite [318].

3.8.2. CRISPR/Cas12 and CRISPR/Cas13-based diagnosis

In 2018, it was found that one of the CRISPR enzymes — Cas12 commenced non-specific cleavage of the single-strand DNA after it had recognized its target DNA. This property of Cas12 can be used to indicate the presence of a specific target, for example, a viral or bacterial genome. Researchers used this discovery to create a technological platform for the detection of nucleic acids, known as DETECTR (DNA Endonuclease Targeted CRISPR Trans Reporter). The platform combines the Cas12a nuclease, its guide RNA specific to the nucleic acid, and a fluorescent reporter molecule. The DETECTR technology was first used to identify and genotype the human papillomavirus (HPV). It took DETECTR one hour to differentiate between HPV16 and HPV18 in the unpurified DNA extracts from human cultured cells and clinical samples. DETECTR correctly (comparable with the result of the PCR test) identified HPV16 in 25 and HPV18 in 23 out of 25 clinical samples [319].

Another important CRISPR/Cas application is the identification of pathogens and detection of specific bacterial genes by using the SHERLOCK (specific high-sensitivity enzymatic reporter unlocking) platform. The platform combines the Cas13a nuclease, its guide RNA specific to the nucleic acid, and a fluorescent reporter molecule. The Cas13a complex binds and cleaves the pre-amplified nucleic acid with high specificity. Using SHERLOCK, it became possible to differentiate between closely related Zika and Dengue virus strains, to genotype several *Escherichia coli* and *Pseudomonas aeruginosa* strains with low cross-reactivity. In addition, the SHERLOCK platform can be used to distinguish between clinical isolates of *Klebsiella pneumoniae* with two different antibiotic resistance genes, thus opening

vast opportunities for multiplex systems combining the identification of bacterial pathogens and detection of antibiotic resistance genes [144].

SHERLOCK was further improved (SHERLOCKv2) to detect up to 4 targets in the same assay. Multiplexing was achieved by combining multiple Cas13 nucleases and a Cas12 nuclease with specific fluorescent reporter complexes, which provided signal detection at different wavelengths. Quantitative detection was achieved by optimizing concentrations of oligonucleotides used during pre-amplification so that the input signal and signal intensity would closely correlate across a broad range of sample concentrations. The enhanced sensitivity was achieved by adding Csm6 to increase the intensity of the cleavage of a fluorescent reporter. It should be noted that SHERLOCKv2 is a portable assay, as the fluorescence readout is replaced with visual detection in lateral flow assays [320].

SHERLOCK can be combined with the HUDSON (heating unextracted diagnostic samples to obliterate nucleases) method, which eliminates the need for nucleic acid extraction and makes it possible to detect pathogens directly in patients' biological specimens (blood, serum or blood plasma, blood cells, saliva, sputum, lymphoid tissues, tissues of hematopoietic organs, and other biological materials). In HUDSON, heating and chemical reduction inactivate nucleases present at high levels in patients' biological specimens; then, viral particles are lysed, releasing nucleic acids into the solution. HUDSON offers high-sensitive detection of the Dengue virus in patients' whole blood, serum, and saliva within 2 hours. HUDSON also makes it possible to distinguish between the four Dengue virus serotypes and detect the 6 most common HIV reverse transcriptase mutations [321].

Thus, CRISPR/Cas opens up promising vistas for future diagnostic systems, including multiplex systems, to be used for identification of bacterial or viral pathogens, for detection of antibiotic resistance genes in bacterial pathogens, and differentiation of closely related strains/isolates of bacterial and viral pathogens.

3.9. Resources for CRISPR/Cas

The CRISPR/Cas system was adapted as a powerful tool for genome editing and became widely used in genomic studies due to its simplicity and cost-effectiveness. The efficiency of the CRISPR/Cas system depends on the properly designed single-guide RNA (sgRNA); therefore, multiple bioinformatics tools were offered to create highly active and specific sgRNA. These tools vary in design, parameters, reference genomes, *etc*. Some of the bioinformatics tools are given in **Table 3.11**, while others can be found at BioinfoGP[5].

[5]URL: https://bioinfogp.cnb.csic.es/tools/wereview/crisprtools

Table 3.11. Bioinformatics tools and resources for the CRISPR/Cas system

No	Resource name	Link	Application	Reference
The design of guide RNAs for the CRISPR/Cas system				
1	CRISPR Guide RNA Design	https://www.benchling.com/crispr/	Guide RNAs can be designed with an imported target sequence; gene coordinates can be specified within the selected genome to be further automatically annotated by Benchling (with the exon and coding sequence information required to design guide RNAs). Sequence import directly from more than 160 reference is possible. With Benchling and the latest algorithms, the designed guide RNAs can be scored by specificity and efficiency to provide selection of the best guides	–
2	GPP sgRNA Designer	https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design	The tool ranks and picks candidate guide RNA sequences for selected target genes, while attempting to maximize on-target effectiveness and minimize off-target activity	[322–325]
3	CHOPCHOP	http://chopchop.cbu.uib.no	The tool ranks and picks candidate guide RNA sequences for selected target genes, analyzing their effectiveness	[326–328]
4	CRISPOR	http://crispor.tefor.net	The program helps design, evaluate and clone sequences coding guide RNAs for the CRISPR/Cas9 system	[329]
5	E-CRISP	http://www.e-crisp.org/E-CRISP/designcrispr.html	The tool identifies candidate guide RNA sequences for selected target genes, while providing information about their estimated effectiveness and off-target activity	[330]

Table 3.11 to be continued on page 102.

No	Resource name	Link	Application	Reference
Prediction and analysis of non-specific activity of the CRISPR/Cas system				
6	Cas-OFFinder	http://www.rgenome.net/cas-offinder/	The fast and universal algorithm for searching potential off-target sites of RNA-guided Cas9 endonucleases (non-specific activity sites).	[331]
7	COSMID	https://crispr.bme.gatech.edu/	The tool for identification and verification of off-target editing sites (non-specific activity)	[332]
8	CRISPR-offinder-v1-2	https://sourceforge.net/projects/crispr-offinder-v1-2/	Taking into account specified target sites and based on the reference genome, this autonomous tool will identify assumed sited of off-target activity and will set the predicted activity based on modeling	[333]
Search for new elements of the CRISPR/Cas system				
9	CRISPRCasFinder	https://crisprcas.i2bc.paris-saclay.fr/CrisprCasFinder/Index	The CRISPRCasFinder program facilitates detection of CRISPR loci and *cas* genes in the sequences provided by the user	[334–338]
10	CRISPRTarget	http://crispr.otago.ac.nz/CRISPRTarget/crispr_analysis.html	The program for identification of the detected CRISPR/Cas systems (bacteriophage and plasmid sequences)	[339]
11	CRISPRmap	http://rna.informatik.uni-freiburg.de/CRISPRmap/Input.jsp	The program provides a quick and detailed insight into CRISPR repeats both in bacterial and archaeal systems. It includes the largest set of CRISPR data and makes it possible to conduct a comprehensive independent clustering analysis to identify families of conserved sequences, potential structure motifs and evolutionary relationships	[340-342]

Table 3.11 to be continued on page 103.

No	Resource name	Link	Application	Reference
12	CRISPRviz	https://github.com/CRISPRlab/CRISPRviz	The program identifies and extracts repeats and spacers from genome files and then displays this information via local web server for additional manipulations	[343]
13	CRISPRStudio	https://github.com/moineaulab/CRISPRStudio	The program was designed to facilitate and accelerate CRISPR array visualization in the analyzed sequences (genomes)	[344]
Post-experimental analysis				
14	CRISPResso2	https://crispresso.pinellolab.partners.org	Analysis of genome editing outcomes by using deep sequencing data (on-target and off-target effects)	[345]
15	Cas-Analyzer	http://www.rgenome.net/cas-analyzer/#!	A JavaScript-based instant assessment tool for high-throughput sequencing data for genomes/genome fragments of edited cells	[346]
16	CRISPR Genome Analyzer	http://crispr-ga.net/	Outcome assessment of high-throughput sequencing data of genome editing	[347]
17	TIDE/TIDER	https://tide.nki.nl/	Analysis of genome editing results by the Sanger sequencing	[348, 349]

References

1. Ishino Y., Shinagawa H., Makino K., et al. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J. Bacteriol.* 1987; 169(12): 5429-33. doi: 10.1128/jb.169.12.5429-5433.1987
2. Mojica F.J., Juez G., Rodríguez-Valera F. Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI site. *Mol. Microbiol.* 1993; 9(3): 613-21. doi: 10.1111/j.1365-2958.1993.tb01721.x
3. Jansen R., Embden J.D., Gaastra W., et al. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* 2002; 43(6): 1565-75. doi: 10.1046/j.1365-2958.2002.02839.x
4. Bolotin A., Quinquis B., Sorokin A., et al. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology (Reading).* 2005; 151(Pt. 8): 2551-61. doi: 10.1099/mic.0.28048-0
5. Makarova K.S., Grishin N.V., Shabalina S.A., et al. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct.* 2006; 1: 7. doi: 10.1186/1745-6150-1-7

6. Mojica F.J., Díez-Villaseñor C., García-Martínez J., et al. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol*. 2005; 60(2): 174-82. doi: 10.1007/s00239-004-0046-3
7. Pourcel C., Salvignol G., Vergnaud G. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology (Reading)*. 2005; 151(Pt. 3): 653-63. doi: 10.1099/mic.0.27437-0
8. Barrangou R., Fremaux C., Deveau H., et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science*. 2007; 315(5819): 1709-12. doi: 10.1126/science.1138140
9. Marraffini L.A., Sontheimer E.J. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science*. 2008; 322(5909): 1843-5. doi: 10.1126/science.1165771
10. Brouns S.J., Jore M.M., Lundgren M., et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*. 2008; 321(5891): 960-4. doi: 10.1126/science.1159689
11. Hale C.R., Zhao P., Olson S., et al. RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell*. 2009; 139(5): 945-56. doi: 10.1016/j.cell.2009.07.040
12. Garneau J.E., Dupuis M.È., Villion M., et al. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*. 2010; 468(7320): 67-71. doi: 10.1038/nature09523
13. Deltcheva E., Chylinski K., Sharma C.M., et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*. 2011; 471(7340): 602-7. doi: 10.1038/nature09886
14. Gasiunas G., Barrangou R., Horvath P., et al. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. USA*. 2012; 109(39): E2579-86. doi: 10.1073/pnas.1208507109
15. Jinek M., Chylinski K., Fonfara I., et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*. 2012; 337(6096): 816-21. doi: 10.1126/science.1225829
16. Cong L., Ran F.A., Cox D., et al. Multiplex genome engineering using CRISPR/Cas systems. *Science*. 2013; 339(6121): 819-23. doi: 10.1126/science.1231143
17. Mali P., Yang L., Esvelt K.M., et al. RNA-guided human genome engineering via Cas9. *Science*. 2013; 339(6121): 823-6. doi: 10.1126/science.1232033
18. Horvath P., Barrangou R. CRISPR/Cas, the immune system of bacteria and archaea. *Science*. 2010; 327(5962): 167-70. doi: 10.1126/science.1179555
19. Wiedenheft B., Sternberg S.H., Doudna J.A. RNA-guided genetic silencing systems in bacteria and archaea. *Nature*. 2012; 482(7385): 331-8. doi: 10.1038/nature10886
20. Miao J., Guo D., Zhang J., et al. Targeted mutagenesis in rice using CRISPR-Cas system. *Cell Res.* 2013; 23(10): 1233-6. doi: 10.1038/cr.2013.123
21. Shan Q., Wang Y., Li J., et al. Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat. Biotechnol.* 2013; 31(8): 686-8. doi: 10.1038/nbt.2650
22. Upadhyay S.K., Kumar J., Alok A., et al. RNA-guided genome editing for target gene mutations in wheat. *G3 (Bethesda)*. 2013; 3(12): 2233-8. doi: 10.1534/g3.113.008847
23. Jiang W., Zhou H., Bi H., et al. Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. *Nucleic Acids Res.* 2013; 41(20): e188. doi: 10.1093/nar/gkt780
24. Liang P., Xu Y., Zhang X., et al. CRISPR/Cas9-mediated gene editing in human tripronuclear zygotes. *Protein Cell.* 2015; 6(5): 363-72. doi: 10.1007/s13238-015-0153-5
25. Anzalone A.V., Randolph P.B., Davis J.R., et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*. 2019; 576(7785): 149-57. doi: 10.1038/s41586-019-1711-4

26. Jackson S.A., McKenzie R.E., Fagerlund R.D., et al. CRISPR-Cas: Adapting to change. *Science*. 2017; 356(6333): eaal5056. doi: 10.1126/science.aal5056
27. Tamulaitis G., Venclovas Č., Siksnys V. Type III CRISPR-Cas immunity: major differences brushed aside. *Trends Microbiol.* 2017; 25(1): 49-61. doi: 10.1016/j.tim.2016.09.012
28. Mohanraju P., Makarova K.S., Zetsche B., et al. Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science.* 2016; 353(6299): aad5147. doi: 10.1126/science.aad5147
29. Charpentier E., Richter H., van der Oost J., et al. Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity. *FEMS Microbiol. Rev.* 2015; 39(3): 428-41. doi: 10.1093/femsre/fuv023
30. Xiao Y., Ng S., Nam K.H., et al. How type II CRISPR-Cas establish immunity through Cas1-Cas2-mediated spacer integration. *Nature.* 2017; 550(7674): 137-41. doi: 10.1038/nature24020
31. He Y., Wang M., Liu M., et al. Cas1 and Cas2 from the type II-C CRISPR-Cas system of Riemerella anatipestifer are required for spacer acquisition. *Front. Cell Infect. Microbiol.* 2018; 8: 195. doi: 10.3389/fcimb.2018.00195
32. Dolan A.E., Hou Z., Xiao Y., et al. Introducing a spectrum of long-range genomic deletions in human embryonic stem cells using type I CRISPR-Cas. *Mol. Cell*. 2019; 74(5): 936-50.e5. doi: 10.1016/j.molcel.2019.03.014
33. Zhang Z., Pan S., Liu T., et al. Cas4 nucleases can effect specific integration of CRISPR spacers. *J. Bacteriol.* 2019; 201(12): e00747-18. doi: 10.1128/JB.00747-18
34. Brendel J., Stoll B., Lange S.J., et al. A complex of Cas proteins 5, 6, and 7 is required for the biogenesis and stability of clustered regularly interspaced short palindromic repeats (crispr)-derived rnas (crrnas) in Haloferax volcanii. *J. Biol. Chem.* 2014; 289(10): 7164-77. doi: 10.1074/jbc.M113.508184
35. Kalwani P., Rath D., Ballal A. Novel molecular aspects of the CRISPR backbone protein 'Cas7' from cyanobacteria. *Biochem. J.* 2020; 477(5): 971-83. doi: 10.1042/BCJ20200026
36. Cass S.D., Haas K.A., Stoll Bet al. The role of Cas8 in type I CRISPR interference. *Biosci. Rep.* 2015; 35(3): e00197. doi: 10.1042/BSR20150043
37. Nayeemul Bari S.M., Hatoum-Aslan A. CRISPR-Cas10 assisted editing of virulent staphylococcal phages. *Methods Enzymol.* 2019; 616: 385-409. doi: 10.1016/bs.mie.2018.10.023
38. Shmakov S.A., Makarova K.S., Wolf Y.I., et al. Systematic prediction of genes functionally linked to CRISPR-Cas systems by gene neighborhood analysis. *Proc. Natl. Acad. Sci. USA.* 2018; 115(23): E5307-16. doi: 10.1073/pnas.1803440115
39. Strohkendl I., Saifuddin F.A., Rybarski J.R., et al. Kinetic basis for DNA target specificity of CRISPR-Cas12a. *Mol. Cell.* 2018; 71(5): 816-24.e3. doi: 10.1016/j.molcel.2018.06.043
40. Cox D.B.T., Gootenberg J.S., Abudayyeh O.O., et al. RNA editing with CRISPR-Cas13. *Science*. 2017; 358(6366): 1019-27. doi: 10.1126/science.aaq0180
41. Aquino-Jarquin G. CRISPR-Cas14 is now part of the artillery for gene editing and molecular diagnostic. *Nanomedicine*. 2019; 18: 428-31. doi: 10.1016/j.nano.2019.03.006
42. Koonin E.V., Makarova K.S. Origins and evolution of CRISPR-Cas systems. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 2019; 374(1772): 20180087. doi: 10.1098/rstb.2018.0087
43. Burstein D., Sun C.L., Brown C.T., et al. Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems. *Nat. Commun.* 2016; 7: 10613. doi: 10.1038/ncomms10613
44. Shmakov S., Smargon A., Scott D., et al. Diversity and evolution of class 2 CRISPR-Cas systems. *Nat. Rev. Microbiol.* 2017; 15(3): 169-82. doi: 10.1038/nrmicro.2016.184

45. Burstein D., Harrington L.B., Strutt S.C., et al. New CRISPR-Cas systems from uncultivated microbes. *Nature.* 2017; 542(7640): 237-41. doi: 10.1038/nature21059

46. Rouillon C., Zhou M., Zhang J., et al. Structure of the CRISPR interference complex CSM reveals key similarities with cascade. *Mol. Cell.* 2013; 52(1): 124-34. doi: 10.1016/j.molcel.2013.08.020

47. Jackson R.N., Wiedenheft B. A conserved structural chassis for mounting versatile CRISPR RNA-guided immune responses. *Mol. Cell.* 2015; 58(5): 722-8. doi: 10.1016/j.molcel.2015.05.023

48. Makarova K.S., Haft D.H., Barrangou R., et al. Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.* 2011; 9(6): 467-77. doi: 10.1038/nrmicro2577

49. Koonin E.V., Makarova K.S. CRISPR-Cas: evolution of an RNA-based adaptive immunity system in prokaryotes. *RNA Biol.* 2013; 10(5): 679-86. doi: 10.4161/rna.24022

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

51. Venclovas Č. Structure of Csm2 elucidates the relationship between small subunits of CRISPR-Cas effector complexes. *FEBS Lett.* 2016; 590(10): 1521-9. doi: 10.1002/1873-3468.12179

52. Makarova K.S., Wolf Y.I., Alkhnbashi O.S., et al. An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* 2015; 13(11): 722-36. doi: 10.1038/nrmicro3569

53. Garrett R.A., Vestergaard G., Shah S.A. Archaeal CRISPR-based immune systems: exchangeable functional modules. *Trends Microbiol.* 2011; 19(11): 549-56. doi: 10.1016/j.tim.2011.08.002

54. Makarova K.S., Wolf Y.I., Iranzo J., et al. Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* 2020; 18(2): 67-83. doi: 10.1038/s41579-019-0299-x

55. Nishimasu H., Ran F.A., Hsu P.D., et al. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell.* 2014; 156(5): 935-49. doi: 10.1016/j.cell.2014.02.001

56. Esvelt K.M., Mali P., Braff J.L., et al. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat. Methods.* 2013; 10(11): 1116-21. doi: 10.1038/nmeth.2681

57. Nishimasu H., Cong L., Yan W.X., et al. Crystal structure of Staphylococcus aureus Cas9. *Cell*. 2015; 162(5): 1113-26. doi: 10.1016/j.cell.2015.08.007

58. Hirano H., Gootenberg J.S., Horii T., et al. Structure and engineering of Francisella novicida Cas9. *Cell*. 2016; 164(5): 950-61. doi: 10.1016/j.cell.2016.01.039

59. Yamada M., Watanabe Y., Gootenberg J.S., et al. Crystal structure of the minimal Cas9 from Campylobacter jejuni reveals the molecular diversity in the CRISPR-Cas9 systems. *Mol. Cell.* 2017; 65(6): 1109-21.e3. doi: 10.1016/j.molcel.2017.02.007

60. Ma H., Naseri A., Reyes-Gutierrez P., et al. Multicolor CRISPR labeling of chromosomal loci in human cells. *Proc. Natl. Acad. Sci. USA.* 2015; 112(10): 3002-7. doi: 10.1073/pnas.1420024112

61. Mojica F.J.M., Díez-Villaseñor C., García-Martínez J., et al. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology.* 2009; 155(Pt. 3): 733-40. doi: 10.1099/mic.0.023960-0

62. Horvath P., Romero D.A., Coûté-Monvoisin A.C., et al. Diversity, activity, and evolution of CRISPR loci in Streptococcus thermophiles. *J. Bacteriol.* 2008; 190(4): 1401-12. doi: 10.1128/JB.01415-07

63. Fonfara I., Le Rhun A., Chylinski K., et al. Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. *Nucleic Acids Res.* 2014; 42(4): 2577-90. doi: 10.1093/nar/gkt1074

64. van der Ploeg J.R. Analysis of CRISPR in Streptococcus mutans suggests frequent occurrence of acquired immunity against infection by M102-like bacteriophages. *Microbiology.* 2009; 155(Pt. 6): 1966-1976. doi: 10.1099/mic.0.027508-0

65. Zhang Y., Heidrich N., Ampattu B.J., et al. Processing-independent CRISPR RNAs limit natural transformation in Neisseria meningitides. *Mol. Cell.* 2013; 50(4): 488-503. doi: 10.1016/j.molcel.2013.05.001

66. Chen H., Choi J., Bailey S. Cut site selection by the two nuclease domains of the Cas9 RNA-guided endonuclease. *J. Biol. Chem.* 2014; 289(19): 13284-94. doi: 10.1074/jbc.M113.539726

67. Mendoza B.J., Trinh C.T. In silico processing of the complete CRISPR-Cas spacer space for identification of PAM sequences. *Biotechnol. J.* 2018; 13(9): e1700595. doi: 10.1002/biot.201700595

68. Hirano S., Abudayyeh O.O., Gootenberg J.S., et al. Structural basis for the promiscuous PAM recognition by Corynebacterium diphtheriae Cas9. *Nat. Commun.* 2019; 10(1): 1968. doi: 10.1038/s41467-019-09741-6

69. Ding Y., Li H., Chen L.L., et al. Recent advances in genome editing using CRISPR/Cas9. *Front Plant Sci*. 2016; 7: 703. doi: 10.3389/fpls.2016.00703

70. Cradick T.J., Fine E.J., Antico C.J., et al. CRISPR/Cas9 systems targeting β-globin and CCR5 genes have substantial off-target activity. *Nucleic Acids Res.* 2013; 41(20): 9584-92. doi: 10.1093/nar/gkt714

71. Fu Y., Foden J.A., Khayter C., et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* 2013; 31(9): 822-6. doi: 10.1038/nbt.2623

72. Hsu P.D., Scott D.A., Weinstein J.A., et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* 2013; 31(9): 827-32. doi: 10.1038/nbt.2647

73. Mali P., Aach J., Stranges P.B., et al. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* 2013; 31(9): 833-8. doi: 10.1038/nbt.2675

74. Pattanayak V., Lin S., Guilinger J.P., et al. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.* 2013; 31(9): 839-43. doi: 10.1038/nbt.2673

75. Tsai S.Q., Zheng Z., Nguyen N.T., et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* 2015; 33(2): 187-97. doi: 10.1038/nbt.3117

76. Kim D., Bae S., Park J., et al. Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat. Methods.* 2015; 12(3): 237-43. doi: 10.1038/nmeth.3284

77. Kuscu C., Arslan S., Singh R., et al. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat. Biotechnol.* 2014; 32(7): 677-83. doi: 10.1038/nbt.2916

78. Ran F.A., Hsu P.D., Lin C.Y., et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell.* 2013; 154(6): 1380-9. doi: 10.1016/j.cell.2013.08.021

79. Guilinger J.P., Thompson D.B., Liu D.R. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat. Biotechnol.* 2014; 32(6): 577-82. doi: 10.1038/nbt.2909

80. Tsai S.Q., Wyvekens N., Khayter C., et al. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat. Biotechnol.* 2014; 32(6): 569-76. doi: 10.1038/nbt.2908

81. Bolukbasi M.F., Gupta A., Oikemus S., et al. DNA-binding-domain fusions enhance the targeting range and precision of Cas9. *Nat. Methods.* 2015; 12(12): 1150-6. doi: 10.1038/nmeth.3624
82. Fu Y., Sander J.D., Reyon D., et al. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat. Biotechnol.* 2014; 32(3): 279-284. doi: 10.1038/nbt.2808
83. Kleinstiver B.P., Pattanayak V., Prew M.S., et al. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature.* 2016; 529(7587): 490-5. doi: 10.1038/nature16526
84. Hwang W.Y., Fu Y., Reyon D., et al. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat. Biotechnol.* 2013; 31(3): 227-9. doi: 10.1038/nbt.2501
85. Ran F.A., Hsu P.D., Wright J., et al. Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* 2013; 8(11): 2281-308. doi: 10.1038/nprot.2013.143
86. Kleinstiver B.P., Prew M.S., Tsai S.Q., et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature.* 2015; 523(7561): 481-5. doi: 10.1038/nature14592
87. Davis K.M., Pattanayak V., Thompson D.B., et al. Small molecule-triggered Cas9 protein with improved genome-editing specificity. *Nat. Chem. Biol.* 2015; 11(5): 316-8. doi: 10.1038/nchembio.1793
88. Slaymaker I.M., Gao L., Zetsche B., et al. Rationally engineered Cas9 nucleases with improved specificity. *Science*. 2016; 351(6268): 84-8. doi: 10.1126/science.aad5227
89. Trevino A.E., Zhang F. Genome editing using Cas9 nickases. *Methods Enzymol.* 2014; 546: 161-74. doi: 10.1016/B978-0-12-801185-0.00008-8
90. Qi L.S., Larson M.H., Gilbert L.A., et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*. 2013; 152(5): 1173-83. doi: 10.1016/j.cell.2013.02.022
91. La Russa M.F., Qi L.S. The new state of the art: Cas9 for gene activation and repression. *Mol. Cell Biol.* 2015; 35(22): 3800-9. doi: 10.1128/MCB.00512-15
92. Konermann S., Brigham M.D., Trevino A.E., et al. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature*. 2015; 517(7536): 583-8. doi: 10.1038/nature14136
93. Jusiak B., Cleto S., Perez-Piñera P., et al. Engineering Synthetic Gene Circuits in Living Cells with CRISPR Technology. *Trends Biotechnol.* 2016; 34(7): 535-47. doi: 10.1016/j.tibtech.2015.12.014
94. Chavez A., Tuttle M., Pruitt B.W., et al. Comparison of Cas9 activators in multiple species. *Nat. Methods.* 2016; 13(7): 563-7. doi: 10.1038/nmeth.3871
95. Zhang Y., Yin C., Zhang T., et al. CRISPR/gRNA-directed synergistic activation mediator (SAM) induces specific, persistent and robust reactivation of the HIV-1 latent reservoirs. *Sci. Rep.* 2015; 5: 16277. doi: 10.1038/srep16277
96. Bialek J.K., Dunay G.A., Voges M., et al. Targeted HIV-1 latency reversal using CRISPR/Cas9-derived transcriptional activator systems. *PLoS One.* 2016; 11(6): e0158294. doi: 10.1371/journal.pone.0158294
97. Chavez A., Scheiman J., Vora S., et al. Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods.* 2015; 12(4): 326-8. doi: 10.1038/nmeth.3312
98. Gilbert L.A., Larson M.H., Morsut L., et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*. 2013; 154(2): 442-51. doi: 10.1016/j.cell.2013.06.044
99. Gilbert L.A., Horlbeck M.A., Adamson B., et al. Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell*. 2014; 159(3): 647-61. doi: 10.1016/j.cell.2014.09.029
100. Shalem O., Sanjana N.E., Zhang F. High-throughput functional genomics using CRISPR-Cas9. *Nat. Rev. Genet.* 2015; 16(5): 299-311. doi: 10.1038/nrg3899

101. Köeferle A., Stricker S.H., Beck S. Brave new epigenomes: the dawn of epigenetic engineering. *Genome Med.* 2015; 7(1): 59. doi: 10.1186/s13073-015-0185-8

102. Hilton I.B., D'Ippolito A.M., Vockley C.M., et al. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat. Biotechnol.* 2015; 33(5): 510-7. doi: 10.1038/nbt.3199

103. Kearns N.A., Pham H., Tabak B., et al. Functional annotation of native enhancers with a Cas9-histone demethylase fusion. *Nat. Methods.* 2015; 12(5): 401-3. doi: 10.1038/nmeth.3325

104. Rudolph T., Beuch S., Reuter G. Lysine-specific histone demethylase LSD1 and the dynamic control of chromatin. *Biol. Chem.* 2013; 394(8): 1019-28. doi: 10.1515/hsz-2013-0119

105. Robertson K.D. DNA methylation and human disease. *Nat. Rev. Genet.* 2005; 6(8): 597-610. doi: 10.1038/nrg1655

106. Xu X., Tao Y., Gao X., et al. A CRISPR-based approach for targeted DNA demethylation. *Cell Discov.* 2016; 2: 16009. doi: 10.1038/celldisc.2016.9

107. Choudhury S.R., Cui Y., Lubecka K., et al. CRISPR-dCas9 mediated TET1 targeting for selective DNA demethylation at BRCA1 promoter. *Oncotarget.* 2016; 7(29): 46545-56. doi: 10.18632/oncotarget.10234

108. Valton J., Dupuy A., Daboussi F., et al. Overcoming transcription activator-like effector (TALE) DNA binding domain sensitivity to cytosine methylation. *J. Biol. Chem.* 2012; 287(46): 38427-32. doi: 10.1074/jbc.C112.408864

109. Vojta A., Dobrinić P., Tadić V., et al. Repurposing the CRISPR-Cas9 system for targeted DNA methylation. *Nucleic Acids Res.* 2016; 44(12): 5615-28. doi: 10.1093/nar/gkw159

110. Ma H., Tu L.C., Naseri A., et al. Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow. *Nat. Biotechnol.* 2016; 34(5): 528-30. doi: 10.1038/nbt.3526

111. Ma H., Tu L.C., Naseri A., et al. CRISPR-Sirius: RNA scaffolds for signal amplification in genome imaging. *Nat. Methods.* 2018; 15(11): 928-31. doi: 10.1038/s41592-018-0174-0

112. Chen B., Gilbert L.A., Cimini B.A., et al. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell.* 2013; 155(7): 1479-91. doi: 10.1016/j.cell.2013.12.001

113. Fujita T., Yuno M., Fujii H. enChIP systems using different CRISPR orthologues and epitope tags. *BMC Res. Notes.* 2018; 11(1): 154. doi: 10.1186/s13104-018-3262-4

114. Fujita T., Yuno M., Fujii H. An enChIP system for the analysis of bacterial genome functions. *BMC Res. Notes.* 2018; 11(1): 387. doi: 10.1186/s13104-018-3486-3

115. Fujita T., Fujii H. Purification of specific DNA species using the CRISPR system. *Biol. Methods Protoc.* 2019; 4(1): bpz008. doi: 10.1093/biomethods/bpz008

116. Fujita T., Kitaura F., Oji A., et al. Transgenic mouse lines expressing the 3xFLAG-dCas9 protein for enChIP analysis. *Genes Cells.* 2018; 23(4): 318-25. doi: 10.1111/gtc.12573

117. Fujita T., Fujii H. Efficient isolation of specific genomic regions and identification of associated proteins by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP) using CRISPR. *Biochem. Biophys. Res. Commun.* 2013; 439(1): 132-6. doi: 10.1016/j.bbrc.2013.08.013

118. Fujita T., Fujii H. Isolation of specific genomic regions and identification of associated molecules by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP) using CRISPR. *Methods Mol. Biol.* 2015; 1288: 43-52. doi: 10.1007/978-1-4939-2474-5_4

119. Schmidtmann E., Anton T., Rombaut P., et al. Determination of local chromatin composition by CasID. *Nucleus*. 2016; 7(5): 476-84. doi: 10.1080/19491034.2016.1239000

120. Tsui C., Inouye C., Levy M., et al. dCas9-targeted locus-specific protein isolation method identifies histone gene regulators. *Proc. Natl. Acad. Sci. USA*. 2018; 115(12): E2734-41. doi: 10.1073/pnas.1718844115

121. Gao X.D., Tu L.C., Mir A., et al. C-BERST: defining subnuclear proteomic landscapes at genomic elements with dCas9-APEX2. *Nat. Methods*. 2018; 15(6): 433-6. doi: 10.1038/s41592-018-0006-2

122. Molla K.A., Yang Y. CRISPR/Cas-mediated base editing: technical considerations and practical applications. *Trends Biotechnol*. 2019; 37(10): 1121-42. doi: 10.1016/j.tibtech.2019.03.008

123. Li X., Wang Y., Liu Y., et al. Base editing with a Cpf1-cytidine deaminase fusion. *Nat. Biotechnol.* 2018; 36(4): 324-7. doi: 10.1038/nbt.4102

124. Yang B., Yang L., Chen J. Development and Application of Base Editors. *CRISPR J.* 2019; 2(2): 91-104. doi: 10.1089/crispr.2019.0001

125. McGrath E., Shin H., Zhang L., et al. Targeting specificity of APOBEC-based cytosine base editor in human iPSCs determined by whole genome sequencing. *Nat. Commun.* 2019; 10(1): 5353. doi: 10.1038/s41467-019-13342-8

126. Xin X., Li J., Zhao D., et al. Double-check base editing for efficient A to G conversions. *ACS Synth. Biol.* 2019; 8(12): 2629-34. doi: 10.1021/acssynbio.9b00284

127. Yang L., Zhang X., Wang L., et al. Increasing targeting scope of adenosine base editors in mouse and rat embryos through fusion of TadA deaminase with Cas9 variants. *Protein Cell*. 2018; 9(9): 814-9. doi: 10.1007/s13238-018-0568-x

128. Li C., Zong Y., Wang Y., et al. Expanded base editing in rice and wheat using a Cas9-adenosine deaminase fusion. *Genome Biol.* 2018; 19(1): 59. doi: 10.1186/s13059-018-1443-z

129. Gaudelli N.M., Lam D.K., Rees H.A., et al. Directed evolution of adenine base editors with increased activity and therapeutic application. *Nat. Biotechnol.* 2020; 38(7): 892-900. doi: 10.1038/s41587-020-0491-6

130. Wu Y., Xu W., Wang F., et al. Increasing cytosine base editing scope and efficiency with engineered Cas9-PmCDA1 fusions and the modified sgRNA in rice. *Front. Genet.* 2019; 10: 379. doi: 10.3389/fgene.2019.00379

131. Komor A.C., Zhao K.T., Packer M.S., et al. 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.* 2017; 3(8): eaao4774. doi: 10.1126/sciadv.aao4774

132. Liu L.D., Huang M., Dai P., et al. Intrinsic nucleotide preference of diversifying base editors guides antibody ex vivo affinity maturation. *Cell Rep*. 2018; 25(4): 884-92.e3. doi: 10.1016/j.celrep.2018.09.090

133. Ledford H. Super-precise new CRISPR tool could tackle a plethora of genetic diseases. *Nature*. 2019; 574(7779): 464-5. doi: 10.1038/d41586-019-03164-5

134. Yamano T., Nishimasu H., Zetsche B., et al. Crystal structure of Cpf1 in complex with guide RNA and target DNA. *Cell*. 2016; 165(4): 949-62. doi: 10.1016/j.cell.2016.04.003

135. Zetsche B., Gootenberg J.S., Abudayyeh O.O., et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*. 2015; 163(3): 759-71. doi: 10.1016/j.cell.2015.09.038

136. Kim D., Kim J., Hur J.K., et al. Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nat. Biotechnol.* 2016; 34(8): 863-8. doi: 10.1038/nbt.3609

137. Ma S., Liu Y., Liu Y., et al. An integrated CRISPR Bombyx mori genome editing system with improved efficiency and expanded target sites. *Insect. Biochem. Mol. Biol.* 2017; 83: 13-20. doi: 10.1016/j.ibmb.2017.02.003

138. Endo A., Masafumi M., Kaya H., et al. Efficient targeted mutagenesis of rice and tobacco genomes using Cpf1 from Francisella novicida. *Sci. Rep.* 2016; 6: 38169. doi: 10.1038/srep38169

139. Zetsche B., Strecker J., Abudayyeh O.O., et al. A survey of genome editing activity for 16 Cas12a orthologs. *Keio J. Med.* 2019; 10.2302/kjm.2019-0009-OA. doi: 10.2302/kjm.2019-0009-OA

140. Teng F., Li J., Cui T., et al. Enhanced mammalian genome editing by new Cas12a orthologs with optimized crRNA scaffolds. *Genome Biol.* 2019; 20(1): 15. doi: 10.1186/s13059-019-1620-8

141. Jacobsen T., Ttofali F., Liao C., et al. Characterization of Cas12a nucleases reveals diverse PAM profiles between closely-related orthologs. *Nucleic Acids Res.* 2020; 48(10): 5624-38. doi: 10.1093/nar/gkaa272

142. Jiang W., Samai P., Marraffini L.A. Degradation of phage transcripts by CRISPR-associated RNases enables type III CRISPR-Cas immunity. *Cell*. 2016; 164(4): 710-21. doi: 10.1016/j.cell.2015.12.053

143. Abudayyeh O.O., Gootenberg J.S., Konermann S., et al. C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science*. 2016; 353(6299): aaf5573. doi: 10.1126/science.aaf5573

144. Gootenberg J.S., Abudayyeh O.O., Lee J.W., et al. Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science*. 2017; 356(6336): 438-42. doi: 10.1126/science.aam9321

145. Tambe A., East-Seletsky A., Knott G.J., et al. RNA binding and HEPN-nuclease activation are decoupled in CRISPR-Cas13a. *Cell Rep.* 2018; 24(4): 1025-36. doi: 10.1016/j.celrep.2018.06.105

146. Konermann S., Lotfy P., Brideau N.J., et al. Transcriptome engineering with RNA-targeting type VI-D CRISPR effectors. *Cell*. 2018; 173(3): 665-76.e14. doi: 10.1016/j.cell.2018.02.033

147. Yan W.X., Chong S., Zhang H., et al. Cas13d Is a compact RNA-targeting type VI CRISPR effector positively modulated by a WYL-domain-containing accessory protein. *Mol. Cell.* 2018; 70(2): 327-39.e5. doi: 10.1016/j.molcel.2018.02.028

148. Harrington L.B., Burstein D., Chen J.S., et al. Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. *Science*. 2018; 362(6416): 839-42. doi: 10.1126/science.aav4294

149. Nemudryi A.A., Valetdinova K.R., Medvedev S.P., et al. TALEN and CRISPR/Cas genome editing systems: tools of discovery. *Acta Naturae.* 2014; 6(3): 19-40.

150. German D.M., Mitalipov S., Mishra A., et al. Therapeutic genome editing in cardiovascular diseases. *JACC Basic Transl. Sci.* 2019; 4(1): 122-31. doi: 10.1016/j.jacbts.2018.11.004

151. Kwarteng A., Ahuno S.T., Kwakye-Nuako G. The therapeutic landscape of HIV-1 via genome editing. *AIDS Res Ther.* 2017; 14(1): 32. doi: 10.1186/s12981-017-0157-8

152. Loureiro A., da Silva G.J. CRISPR-Cas: Converting a bacterial defence mechanism into a state-of-the-art genetic manipulation tool. *Antibiotics (Basel).* 2019; 8(1): 18. doi: 10.3390/antibiotics8010018

153. Karimian A., Azizian K., Parsian H., et al. CRISPR/Cas9 technology as a potent molecular tool for gene therapy. *J. Cell Physiol.* 2019; 234(8): 12267-77. doi: 10.1002/jcp.27972

154. Caplice E., Fitzgerald G.F. Food fermentations: role of microorganisms in food production and preservation. *Int. J. Food Microbiol.* 1999; 50(1-2): 131-49. doi: 10.1016/s0168-1605(99)00082-3

155. Hill C., Guarner F., Reid G., et al. Expert consensus document. The International Scientific Association for Probiotics and Prebiotics consensus statement on the scope and appropriate use of the term probiotic. *Nat. Rev. Gastroenterol. Hepatol.* 2014; 11(8): 506-14. doi: 10.1038/nrgastro.2014.66

156. Briner A.E., Lugli G.A., Milani C., et al. Occurrence and diversity of CRISPR-Cas systems in the genus bifidobacterium. *PLoS One.* 2015; 10(7): e0133661. doi: 10.1371/journal.pone.0133661

157. Sun Z., Harris H.M., McCann A., et al. Expanding the biotechnology potential of lactobacilli through comparative genomics of 213 strains and associated genera. *Nat. Commun.* 2015; 6: 8322. doi: 10.1038/ncomms9322

158. Arslan Z., Hermanns V., Wurm R., et al. Detection and characterization of spacer integration intermediates in type I-E CRISPR-Cas system. *Nucleic Acids Res.* 2014; 42(12): 7884-93. doi: 10.1093/nar/gku510

159. Barrangou R., Coûté-Monvoisin A.C., Stahl B., et al. Genomic impact of CRISPR immunization against bacteriophages. *Biochem. Soc. Trans.* 2013; 41(6): 1383-91. doi: 10.1042/BST20130160

160. Andersson A.F., Banfield J.F. Virus population dynamics and acquired virus resistance in natural microbial communities. *Science*. 2008; 320(5879): 1047-50. doi: 10.1126/science.1157358

161. Tyson G.W., Banfield J.F. Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. *Environ. Microbiol.* 2008; 10(1): 200-7. doi: 10.1111/j.1462-2920.2007.01444.x

162. Barrangou R., Horvath P. CRISPR: new horizons in phage resistance and strain identification. *Annu. Rev. Food Sci. Technol.* 2012; 3: 143-62. doi: 10.1146/annurev-food-022811-101134

163. Barrangou R., Dudley E.G. CRISPR-based typing and next-generation tracking technologies. *Annu. Rev. Food Sci. Technol.* 2016; 7: 395-411. doi: 10.1146/annurev-food-022814-015729

164. Selle K., Barrangou R. CRISPR-based technologies and the future of food science. *J. Food Sci.* 2015; 80(11): R2367-72. doi: 10.1111/1750-3841.13094

165. Broadbent J.R., Neeno-Eckwall E.C., Stahl B., et al. Analysis of the Lactobacillus casei supragenome and its influence in species evolution and lifestyle adaptation. *BMC Genomics.* 2012; 13: 533. doi: 10.1186/1471-2164-13-533

166. Smokvina T., Wels M., Polka J., et al. Lactobacillus paracasei comparative genomics: towards species pan-genome definition and exploitation of diversity. *PLoS One.* 2013; 8(7): e68731. doi: 10.1371/journal.pone.0068731

167. Douillard F.P., Ribbera A., Kant R., et al. Comparative genomic and functional analysis of 100 Lactobacillus rhamnosus strains and their comparison with strain GG. *PLoS Genet.* 2013; 9(8): e1003683. doi: 10.1371/journal.pgen.1003683

168. Hullahalli K., Rodrigues M., Schmidt B.D., et al. Comparative analysis of the orphan CRISPR2 locus in 242 Enterococcus faecalis strains. *PLoS One.* 2015; 10(9): e0138890. doi: 10.1371/journal.pone.0138890

169. Sanozky-Dawes R., Selle K., O'Flaherty S., et al. Occurrence and activity of a type II CRISPR-Cas system in Lactobacillus gasseri. *Microbiology.* 2015; 161(9): 1752-61. doi: 10.1099/mic.0.000129

170. Briner A.E., Barrangou R. Lactobacillus buchneri genotyping on the basis of clustered regularly interspaced short palindromic repeat (CRISPR) locus diversity. *Appl. Environ. Microbiol.* 2014; 80(3): 994-1001. doi: 10.1128/AEM.03015-13

171. DiMarzio M., Shariat N., Kariyawasam S., et al. Antibiotic resistance in Salmonella enterica serovar typhimurium associates with CRISPR sequence type. *Antimicrob. Agents Chemother.* 2013; 57(9): 4282-9. doi: 10.1128/AAC.00913-13

172. Fabre L., Zhang J., Guigon G., et al. CRISPR typing and subtyping for improved laboratory surveillance of Salmonella infections. *PLoS One.* 2012; 7(5): e36995. doi: 10.1371/journal.pone.0036995

173. Fricke W.F., Mammel M.K., McDermott P.F., et al. Comparative genomics of 28 Salmonella enterica isolates: evidence for CRISPR-mediated adaptive sublineage evolution. *J. Bacteriol.* 2011; 193(14): 3556-68. doi: 10.1128/JB.00297-11

174. Liu F., Barrangou R., Gerner-Smidt P., et al. Novel virulence gene and clustered regularly interspaced short palindromic repeat (CRISPR) multilocus sequence typing scheme for subtyping of the major serovars of Salmonella enterica subsp. Enterica. *Appl. Environ. Microbiol.* 2011; 77(6): 1946-56. doi: 10.1128/AEM.02625-10

175. Liu F., Kariyawasam S., Jayarao B.M., et al. Subtyping Salmonella enterica serovar enteritidis isolates from different sources by using sequence typing based on virulence genes and clustered regularly interspaced short palindromic repeats (CRISPRs). *Appl. Environ. Microbiol.* 2011; 77(13): 4520-6. doi: 10.1128/AEM.00468-11

176. Pettengill J.B., Timme R.E., Barrangou R., et al. The evolutionary history and diagnostic utility of the CRISPR-Cas system within Salmonella enterica ssp. Enterica. *PeerJ*. 2014; 2: e340. doi: 10.7717/peerj.340

177. Shariat N., Kirchner M.K., Sandt C.H., et al. Subtyping of Salmonella enterica serovar Newport outbreak isolates by CRISPR-MVLST and determination of the relationship between CRISPR-MVLST and PFGE results. *J. Clin. Microbiol.* 2013; 51(7): 2328-36. doi: 10.1128/JCM.00608-13

178. Shariat N., Sandt C.H., DiMarzio M.J., et al. CRISPR-MVLST subtyping of Salmonella enterica subsp. enterica serovars Typhimurium and Heidelberg and application in identifying outbreak isolates. *BMC Microbiol.* 2013; 13: 254. doi: 10.1186/1471-2180-13-254

179. Shariat N., Dudley E.G. CRISPRs: molecular signatures used for pathogen subtyping. *Appl. Environ. Microbiol.* 2014; 80(2): 430-9. doi: 10.1128/AEM.02790-13

180. Shariat N., DiMarzio M.J., Yin S., et al. The combination of CRISPR-MVLST and PFGE provides increased discriminatory power for differentiating human clinical isolates of Salmonella enterica subsp. enterica serovar Enteritidis. *Food Microbiol.* 2013; 34(1): 164-73. doi: 10.1016/j.fm.2012.11.012

181. Timme R.E., Pettengill J.B., Allard M.W., et al. Phylogenetic diversity of the enteric pathogen Salmonella enterica subsp. enterica inferred from genome-wide reference-free SNP characters. *Genome Biol. Evol.* 2013; 5(11): 2109-23. doi: 10.1093/gbe/evt159

182. Wehnes C.A., Rehberger T.G., Barrangou R., et al. Short communication: Determination of Salmonella clustered regularly interspaced short palindromic repeats (CRISPR) diversity on dairy farms in Wisconsin and Minnesota. *J. Dairy Sci.* 2014; 97(10): 6370-7. doi: 10.3168/jds.2013-7595

183. Kovanen S.M., Kivistö R.I., Rossi M., et al. A combination of MLST and CRISPR typing reveals dominant Campylobacter jejuni types in organically farmed laying hens. *J. Appl. Microbiol.* 2014; 117(1): 249-57. doi: 10.1111/jam.12503

184. Hargreaves K.R., Flores C.O., Lawley T.D., et al. Abundant and diverse clustered regularly interspaced short palindromic repeat spacers in Clostridium difficile strains and prophages target multiple phage types within this pathogen. *mBio.* 2014; 5(5): e01045-13. doi: 10.1128/mBio.01045-13

185. Mokrousov I., Vyazovaya A., Kolodkina V., et al. Novel macroarray-based method of Corynebacterium diphtheriae genotyping: evaluation in a field study in Belarus. *Eur. J. Clin. Microbiol. Infect. Dis.* 2009; 28(6): 701-3. doi: 10.1007/s10096-008-0674-4

186. Jiang Y., Yin S., Dudley E.G., et al. Diversity of CRISPR loci and virulence genes in pathogenic Escherichia coli isolates from various sources. *Int. J. Food Microbiol.* 2015; 204: 41-6. doi: 10.1016/j.ijfoodmicro.2015.03.025

187. D'Auria G., Jiménez-Hernández N., Peris-Bondia F., et al. Legionella pneumophila pangenome reveals strain-specific virulence factors. *BMC Genomics*. 2010; 11: 181. doi: 10.1186/1471-2164-11-181

188. Kinnevey P.M., Shore A.C., Brennan G.I., et al. Emergence of sequence type 779 methicillin-resistant Staphylococcus aureus harboring a novel pseudo staphylococcal cassette chromosome mec (SCCmec)-SCC-SCCCRISPR composite element in Irish hospitals. *Antimicrob. Agents Chemother.* 2013; 57(1): 524-31. doi: 10.1128/AAC.01689-12

189. Sun H., Li Y., Shi X., et al. Association of CRISPR/Cas evolution with Vibrio parahaemolyticus virulence factors and genotypes. *Foodborne Pathog. Dis.* 2015; 12(1): 68-73. doi: 10.1089/fpd.2014.1792

190. Groenen P.M., Bunschoten A.E., van Soolingen D., et al. Nature of DNA polymorphism in the direct repeat cluster of Mycobacterium tuberculosis; application for strain differentiation by a novel typing method. *Mol. Microbiol.* 1993; 10(5): 1057-65. doi: 10.1111/j.1365-2958.1993.tb00976.x

191. Kamerbeek J., Schouls L., Kolk A., et al. Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology. *J. Clin. Microbiol.* 1997; 35(4): 907-14.

192. Cui Y., Li Y., Gorgé O., Platonov M.E., et al. Insight into microevolution of Yersinia pestis by clustered regularly interspaced short palindromic repeats. *PLoS One.* 2008; 3(7): e2652. doi: 10.1371/journal.pone.0002652

193. Edmond E., Moineau S. Bacteriophages and food fermentations. In: McGrath S., van Sinderen D., eds. *Bacteriophage: Genetics and Molecular Biology.* Poole, UK: Caister Acad.; 2007: 93-124.

194. Marcó M.B., Moineau S., Quiberoni A. Bacteriophages and dairy fermentations. *Bacteriophage*. 2012; 2(3): 149-58. doi: 10.4161/bact.21868

195. Samson J.E., Moineau S. Bacteriophages in food fermentations: new frontiers in a continuous arms race. *Annu. Rev. Food Sci. Technol.* 2013; 4: 347-68. doi: 10.1146/annurev-food-030212-182541

196. Djordjevic G.M., O'Sullivan D.J., Walker S.A., et al. A triggered-suicide system designed as a defense against bacteriophages. *J. Bacteriol.* 1997; 179(21): 6741-8. doi: 10.1128/jb.179.21.6741-6748.1997

197. Durmaz E., Klaenhammer T.R. Abortive phage resistance mechanism AbiZ speeds the lysis clock to cause premature lysis of phage-infected Lactococcus lactis. *J. Bacteriol.* 2007; 189(4): 1417-25. doi: 10.1128/JB.00904-06

198. Higgins D.L., Sanozky-Dawes R.B., Klaenhammer T.R. Restriction and modification activities from Streptococcus lactis ME2 are encoded by a self-transmissible plasmid, pTN20, that forms cointegrates during mobilization of lactose-fermenting ability. *J. Bacteriol.* 1988; 170(8): 3435-42. doi: 10.1128/jb.170.8.3435-3442.1988

199. Labrie S.J., Samson J.E., Moineau S. Bacteriophage resistance mechanisms. *Nat. Rev. Microbiol.* 2010; 8(5): 317-27. doi: 10.1038/nrmicro2315

200. Sanders M.E., Klaenhammer T.R. Evidence for Plasmid Linkage of Restriction and Modification in Streptococcus cremoris KH. *Appl. Environ. Microbiol.* 1981; 42(6): 944-50.

201. Sanders M.E., Klaenhammer T.R. Characterization of phage-sensitive mutants from a phage-insensitive strain of Streptococcus lactis: Evidence for a plasmid determinant that prevents phage adsorption. *Appl. Environ. Microbiol.* 1983; 46(5): 1125-33.

202. Edgar R., Qimron U. The Escherichia coli CRISPR system protects from λ lysogenization, lysogens, and prophage induction. *J. Bacteriol.* 2010; 192(23): 6291-4. doi: 10.1128/JB.00644-10

203. Nozawa T., Furukawa N., Aikawa C., et al. CRISPR inhibition of prophage acquisition in Streptococcus pyogenes. *PLoS One.* 2011; 6(5): e19543. doi: 10.1371/journal.pone.0019543

204. Palmer K.L., Gilmore M.S. Multidrug-resistant enterococci lack CRISPR-cas. *mBio.* 2010; 1(4): e00227-10. doi: 10.1128/mBio.00227-10

205. Shimomura Y., Okumura K., Murayama S.Y., et al. Complete genome sequencing and analysis of a Lancefield group G Streptococcus dysgalactiae subsp. equisimilis strain causing streptococcal toxic shock syndrome (STSS). *BMC Genomics.* 2011; 12: 17. doi: 10.1186/1471-2164-12-17

206. Bourdichon F., Casaregola S., Farrokh C., et al. Food fermentations: microorganisms with technological beneficial use. *Int. J. Food Microbiol.* 2012; 154(3): 87-97. doi: 10.1016/j.ijfoodmicro.2011.12.030

207. Sapranauskas R., Gasiunas G., Fremaux C., et al. The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. *Nucleic Acids Res.* 2011; 39(21): 9275-82. doi: 10.1093/nar/gkr606

208. Selle K., Barrangou R. Harnessing CRISPR-Cas systems for bacterial genome editing. *Trends Microbiol.* 2015; 23(4): 225-32. doi: 10.1016/j.tim.2015.01.008

209. Jiang W., Bikard D., Cox D., et al. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* 2013; 31(3): 233-9. doi: 10.1038/nbt.2508

210. Jiang Y., Chen B., Duan C., et al. Multigene editing in the Escherichia coli genome via the CRISPR-Cas9 system. *Appl. Environ. Microbiol.* 2015; 81(7): 2506-14. doi: 10.1128/AEM.04023-14

211. Oh J.H., van Pijkeren J.P. CRISPR-Cas9-assisted recombineering in Lactobacillus reuteri. *Nucleic Acids Res.* 2014; 42(17): e131. doi: 10.1093/nar/gku623

212. Charpentier E., Marraffini L.A. Harnessing CRISPR-Cas9 immunity for genetic engineering. *Curr. Opin. Microbiol.* 2014; 19: 114-9. doi: 10.1016/j.mib.2014.07.001

213. Gophna U., Allers T., Marchfelder A. Finally, archaea get their CRISPR-Cas toolbox. *Trends Microbiol*. 2017; 25(6): 430-432. doi: 10.1016/j.tim.2017.03.009

214. Barrangou R., van Pijkeren J.P. Exploiting CRISPR-Cas immune systems for genome editing in bacteria. *Curr. Opin. Biotechnol.* 2016; 37: 61-8. doi: 10.1016/j.copbio.2015.10.003

215. Selle K., Klaenhammer T.R., Barrangou R. CRISPR-based screening of genomic island excision events in bacteria. *Proc. Natl. Acad. Sci. USA*. 2015; 112(26): 8076-81. doi: 10.1073/pnas.1508525112

216. Mougiakos I., Bosma E.F., de Vos W.M., et al. Next generation prokaryotic engineering: The CRISPR-Cas toolkit. *Trends Biotechnol.* 2016; 34(7): 575-87. doi: 10.1016/j.tibtech.2016.02.004

217. Bikard D., Barrangou R. Using CRISPR-Cas systems as antimicrobials. *Curr. Opin. Microbiol.* 2017; 37: 155-60. doi: 10.1016/j.mib.2017.08.005

218. Goren M., Yosef I., Qimron U. Sensitizing pathogens to antibiotics using the CRISPR-Cas system. *Drug Resist. Updat.* 2017; 30: 1-6. doi: 10.1016/j.drup.2016.11.001

219. Gomaa A.A., Klumpe H.E., Luo M.L., et al. Programmable removal of bacterial strains by use of genome-targeting CRISPR-Cas systems. *mBio*. 2014; 5(1): e00928-13. doi: 10.1128/mBio.00928-13

220. Hagens S., Bläsi U. Genetically modified filamentous phage as bactericidal agents: a pilot study. *Lett. Appl. Microbiol.* 2003; 37(4): 318-23. doi: 10.1046/j.1472-765x.2003.01400.x

221. Moradpour Z., Sepehrizadeh Z., Rahbarizadeh F., et al. Genetically engineered phage harbouring the lethal catabolite gene activator protein gene with an inducer-independent promoter for biocontrol of Escherichia coli. *FEMS Microbiol. Lett.* 2009; 296(1): 67-71. doi: 10.1111/j.1574-6968.2009.01620.x

222. Westwater C., Kasman L.M., Schofield D.A., et al. Use of genetically engineered phage to deliver antimicrobial agents to bacteria: an alternative therapy for treatment of bacterial infections. *Antimicrob. Agents Chemother.* 2003; 47(4): 1301-7. doi: 10.1128/aac.47.4.1301-1307.2003

223. Hagens S., Habel A., von Ahsen U., et al. Therapy of experimental pseudomonas infections with a nonreplicating genetically modified phage. *Antimicrob. Agents Chemother.* 2004; 48(10): 3817-22. doi: 10.1128/AAC.48.10.3817-3822.2004

224. Citorik R.J., Mimee M., Lu T.K. Sequence-specific antimicrobials using efficiently delivered RNA-guided nucleases. *Nat. Biotechnol.* 2014; 32(11): 1141-5. doi: 10.1038/nbt.3011

225. Bikard D., Euler C.W., Jiang W., et al. Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials. *Nat. Biotechnol.* 2014; 32(11): 1146-50. doi: 10.1038/nbt.3043

226. Yosef I., Manor M., Kiro R., et al. Temperate and lytic bacteriophages programmed to sensitize and kill antibiotic-resistant bacteria. *Proc. Natl. Acad. Sci. USA*. 2015; 112(23): 7267-72. doi: 10.1073/pnas.1500107112

227. Oakes B.L., Fellmann C., Rishi H., et al. CRISPR-Cas9 circular permutants as programmable scaffolds for genome modification. *Cell.* 2019; 176(1-2): 254-67.e16. doi: 10.1016/j.cell.2018.11.052

228. Yuan J., Ma Y., Huang T., et al. Genetic modulation of RNA splicing with a CRISPR-guided cytidine deaminase. *Mol. Cell.* 2018; 72(2): 380-94.e7. doi: 10.1016/j.molcel.2018.09.002

229. Abudayyeh O.O., Gootenberg J.S., Essletzbichler P., et al. RNA targeting with CRISPR-Cas13. *Nature*. 2017; 550(7675): 280-4. doi: 10.1038/nature24049

230. Xie N., Zhou Y., Sun Q., et al. Novel epigenetic techniques provided by the CRISPR/Cas9 system. *Stem Cells Int.* 2018; 2018: 7834175. doi: 10.1155/2018/7834175

231. Wu S.S., Li Q.C., Yin C.Q., et al. Advances in CRISPR/Cas-based gene therapy in human genetic diseases. *Theranostics*. 2020; 10(10): 4374-82. doi: 10.7150/thno.43360

232. Papasavva P., Kleanthous M., Lederer C.W. Rare opportunities: CRISPR/Cas-based therapy development for rare genetic diseases. *Mol. Diagn. Ther.* 2019; 23(2): 201-22. doi: 10.1007/s40291-019-00392-3

233. Kennedy E.M., Cullen B.R. Gene editing: a new tool for viral disease. *Annu. Rev. Med.* 2017; 68: 401-11. doi: 10.1146/annurev-med-051215-031129

234. Xiong X., Chen M., Lim W.A., et al. CRISPR/Cas9 for human genome engineering and disease research. *Annu. Rev. Genomics Hum. Genet.* 2016; 17: 131-54. doi: 10.1146/annurev-genom-083115-022258.

235. Ferdosi S.R., Ewaisha R., Moghadam F., et al. Multifunctional CRISPR-Cas9 with engineered immunosilenced human T cell epitopes. *Nat. Commun.* 2019; 10(1): 1842. doi: 10.1038/s41467-019-09693-x

236. Li H., Yang Y., Hong W., et al. Applications of genome editing technology in the targeted therapy of human diseases: mechanisms, advances and prospects. *Signal Transduct. Target Ther.* 2020; 5: 1. doi: 10.1038/s41392-019-0089-y

237. Chen S., Yu X., Guo D. CRISPR-Cas Targeting of host genes as an antiviral strategy. *Viruses.* 2018; 10(1): 40. doi: 10.3390/v10010040

238. Doms R.W. Chemokine receptors and HIV entry. *AIDS*. 2001; 15(Suppl. 1): S34-5. doi: 10.1097/00002030-200102001-00051

239. O'Brien S.J., Moore J.P. The effect of genetic variation in chemokines and their receptors on HIV transmission and progression to AIDS. *Immunol. Rev.* 2000; 177: 99-111. doi: 10.1034/j.1600-065x.2000.17710.x

240. Rose J.D., Rhea A.M., Weber J., et al. Current tests to evaluate HIV-1 coreceptor tropism. *Curr. Opin. HIV AIDS*. 2009; 4(2): 136-42. doi: 10.1097/COH.0b013e328322f973

241. Koujah L., Shukla D., Naqvi A.R. CRISPR-Cas based targeting of host and viral genes as an antiviral strategy. *Semin. Cell Dev. Biol.* 2019; 96: 53-64. doi: 10.1016/j.semcdb.2019.04.004

242. Liu Z., Chen S., Jin X., et al. Genome editing of the HIV co-receptors CCR5 and CXCR4 by CRISPR-Cas9 protects $CD4^+$ T cells from HIV-1 infection. *Cell Biosci.* 2017; 7: 47. doi: 10.1186/s13578-017-0174-2

243. Xu L., Yang H., Gao Y., et al. CRISPR/Cas9-Mediated CCR5 ablation in human hematopoietic stem/progenitor cells confers HIV-1 resistance in vivo. *Mol. Ther.* 2017; 25(8): 1782-9. doi: 10.1016/j.ymthe.2017.04.027

244. Li C., Guan X., Du T., et al. Inhibition of HIV-1 infection of primary CD4+ T-cells by gene editing of CCR5 using adenovirus-delivered CRISPR/Cas9. *J. Gen. Virol.* 2015; 96(8): 2381-93. doi: 10.1099/vir.0.000139

245. Cho S.W., Kim S., Kim Y., et al. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res.* 2014; 24(1): 132-41. doi: 10.1101/gr.162339.113

246. Hou P., Chen S., Wang S., et al. Genome editing of CXCR4 by CRISPR/cas9 confers cells resistant to HIV-1 infection. *Sci. Rep.* 2015; 5: 15577. doi: 10.1038/srep15577

247. Schumann K., Lin S., Boyer E., et al. Generation of knock-in primary human T cells using Cas9 ribonucleoproteins. *Proc. Natl. Acad. Sci. USA*. 2015; 112(33): 10437-42. doi: 10.1073/pnas.1512503112

248. Liu S., Wang Q., Yu X., et al. HIV-1 inhibition in cells with CXCR4 mutant genome created by CRISPR-Cas9 and piggyBac recombinant technologies. *Sci. Rep.* 2018; 8(1): 8573. doi: 10.1038/s41598-018-26894-4

249. Strich J.R., Chertow D.S. CRISPR-Cas biology and its application to infectious diseases. *J. Clin. Microbiol.* 2019; 57(4): e01307-18. doi: 10.1128/JCM.01307-18

250. Hu W., Kaminski R., Yang F., et al. RNA-directed gene editing specifically eradicates latent and prevents new HIV-1 infection. *Proc. Natl. Acad. Sci. USA*. 2014; 111(31): 11461-6. doi: 10.1073/pnas.1405186111

251. Wang G., Zhao N., Berkhout B., et al. CRISPR-Cas based antiviral strategies against HIV-1. *Virus Res.* 2018; 244: 321-32. doi: 10.1016/j.virusres.2017.07.020

252. Yin C., Zhang T., Qu X., et al. In vivo excision of HIV-1 provirus by saCas9 and multiplex single-guide RNAs in animal models. *Mol. Ther.* 2017; 25(5): 1168-86. doi: 10.1016/j.ymthe.2017.03.012

253. Bella R., Kaminski R., Mancuso P., et al. Removal of HIV DNA by CRISPR from patient blood engrafts in humanized mice. *Mol. Ther. Nucleic Acids*. 2018; 12: 275-82. doi: 10.1016/j.omtn.2018.05.021

254. World Health Organization; 2018. Finding a cure for hepatitis B: are we close? Available at: http://www.who.int/hepatitis/news-events/hbv-cure-overview/en

255. Li H., Sheng C., Wang S., et al. Removal of integrated hepatitis B virus DNA using CRISPR-Cas9. *Front. Cell Infect. Microbiol.* 2017; 7: 91. doi: 10.3389/fcimb.2017.00091

256. Scott T., Moyo B., Nicholson S., et al. ssAAVs containing cassettes encoding SaCas9 and guides targeting hepatitis B virus inactivate replication of the virus in cultured cells. *Sci. Rep.* 2017; 7(1): 7401. doi: 10.1038/s41598-017-07642-6

257. van Diemen F.R., Kruse E.M., Hooykaas M.J., et al. CRISPR/Cas9-mediated genome editing of herpesviruses limits productive and latent infections. *PLoS Pathog.* 2016; 12(6): e1005701. doi: 10.1371/journal.ppat.1005701

258. Roehm P.C., Shekarabi M., Wollebo H.S., et al. Inhibition of HSV-1 replication by gene editing strategy. *Sci. Rep.* 2016; 6: 23146. doi: 10.1038/srep23146

259. Wollebo H.S., Bellizzi A., Kaminski R., et al. CRISPR/Cas9 system as an agent for eliminating polyomavirus JC infection. *PLoS One.* 2015; 10(9): e0136046. doi: 10.1371/journal.pone.0136046

260. Kennedy E.M., Kornepati A.V., Goldstein M., et al. Inactivation of the human papillomavirus E6 or E7 gene in cervical carcinoma cells by using a bacterial CRISPR/Cas RNA-guided endonuclease. *J. Virol.* 2014; 88(20): 11965-72. doi: 10.1128/JVI.01879-14

261. Rodwell C., Aymé S. 2014 Report on the state of the art of rare disease activities in Europe. Part II: Key developments in the field of rare diseases in Europe in 2013. European Union, Brussels, Belgium; 2014.

262. Porteus M.H. Towards a new era in medicine: therapeutic genome editing. *Genome Biol.* 2015; 16: 286. doi: 10.1186/s13059-015-0859-y

263. Vaidyanathan S., Salahudeen A.A., Sellers Z.M., et al. High-efficiency, selection-free gene repair in airway stem cells from cystic fibrosis patients rescues CFTR function in differentiated epithelia. *Cell Stem Cell.* 2020; 26(2): 161-71.e4. doi: 10.1016/j.stem.2019.11.002

264. Schwank G., Koo B.K., Sasselli V., et al. Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. *Cell Stem Cell.* 2013; 13(6): 653-8. doi: 10.1016/j.stem.2013.11.002

265. Firth A.L., Menon T., Parker G.S., et al. Functional gene correction for cystic fibrosis in lung epithelial cells generated from patient iPSCs. *Cell Rep.* 2015; 12(9): 1385-90. doi: 10.1016/j.celrep.2015.07.062

266. Vakulskas C.A., Dever D.P., Rettig G.R., et al. A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells. *Nat. Med.* 2018; 24(8): 1216-24. doi: 10.1038/s41591-018-0137-0

267. Lomova A., Clark D.N., Campo-Fernandez B., et al. Improving gene editing outcomes in human hematopoietic stem and progenitor cells by temporal control of DNA repair. *Stem Cells*. 2019; 37(2): 284-94. doi: 10.1002/stem.2935

268. Martin R.M., Ikeda K., Cromer M.K., et al. Highly efficient and marker-free genome editing of human pluripotent stem cells by CRISPR-Cas9 RNP and AAV6 donor-mediated homologous recombination. *Cell Stem Cell.* 2019; 24(5): 821-8.e5. doi: 10.1016/j.stem.2019.04.001

269. Wattanapanitch M., Damkham N., Potirat P., et al. One-step genetic correction of hemoglobin E/beta-thalassemia patient-derived iPSCs by the CRISPR/Cas9 system. *Stem Cell Res. Ther.* 2018; 9(1): 46. doi: 10.1186/s13287-018-0779-3

270. Patsali P., Turchiano G., Papasavva P., et al. Correction of IVS I-110(G>A) β-thalassemia by CRISPR/Cas-and TALEN-mediated disruption of aberrant regulatory elements in human hematopoietic stem and progenitor cells. *Haematologica*. 2019; 104(11): e497-501. doi: 10.3324/haematol.2018.215178

271. Niu X., He W., Song B., et al. Combining single strand oligodeoxynucleotides and CRISPR/Cas9 to correct gene mutations in β-thalassemia-induced pluripotent stem cells. *J. Biol. Chem.* 2016; 291(32): 16576-85. doi: 10.1074/jbc.M116.719237

272. Monteys A.M., Ebanks S.A., Keiser M.S., et al. CRISPR/Cas9 editing of the mutant huntingtin allele in vitro and in vivo. *Mol. Ther.* 2017; 25(1): 12-23. doi: 10.1016/j.ymthe.2016.11.010

273. Yang S., Chang R., Yang H., et al. CRISPR/Cas9-mediated gene editing ameliorates neurotoxicity in mouse model of Huntington's disease. *J. Clin. Invest.* 2017; 127(7): 2719-24. doi: 10.1172/JCI92087

274. Merienne N., Vachey G., de Longprez L., et al. The self-inactivating KamiCas9 system for the editing of CNS disease genes. *Cell Rep.* 2017; 20(12): 2980-91. doi: 10.1016/j.celrep.2017.08.075

275. Ekman F.K., Ojala D.S., Adil M.M., et al. CRISPR-Cas9-mediated genome editing increases lifespan and improves motor deficits in a Huntington's disease mouse model. *Mol. Ther. Nucleic Acids.* 2019; 17: 829-39. doi: 10.1016/j.omtn.2019.07.009

276. Nelson C.E., Hakim C.H., Ousterout D.G., et al. In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy. *Science*. 2016; 351(6271): 403-7. doi: 10.1126/science.aad5143

277. Nelson C.E., Wu Y., Gemberling M.P., et al. Long-term evaluation of AAV-CRISPR genome editing for Duchenne muscular dystrophy. *Nat. Med.* 2019; 25(3): 427-32. doi: 10.1038/s41591-019-0344-3

278. Min Y.L., Li H., Rodriguez-Caycedo C., et al. CRISPR-Cas9 corrects Duchenne muscular dystrophy exon 44 deletion mutations in mice and human cells. *Sci. Adv.* 2019; 5(3): eaav4324. doi: 10.1126/sciadv.aav4324

279. Bengtsson N.E., Hall J.K., Odom G.L., et al. Muscle-specific CRISPR/Cas9 dystrophin gene editing ameliorates pathophysiology in a mouse model for Duchenne muscular dystrophy. *Nat. Commun.* 2017; 8: 14454. doi: 10.1038/ncomms14454

280. Amoasii L., Hildyard J.C.W., Li H., et al. Gene editing restores dystrophin expression in a canine model of Duchenne muscular dystrophy. *Science*. 2018; 362(6410): 86-91. doi: 10.1126/science.aau1549

281. Zhang Y., Long C., Li H., et al. CRISPR-Cpf1 correction of muscular dystrophy mutations in human cardiomyocytes and mice. *Sci. Adv.* 2017; 3(4): e1602814. doi: 10.1126/sciadv.1602814

282. Stephens C.J., Lauron E.J., Kashentseva E., et al. Long-term correction of hemophilia B using adenoviral delivery of CRISPR/Cas9. *J. Control. Release.* 2019; 298: 128-41. doi: 10.1016/j.jconrel.2019.02.009

283. Hu Z., Zhou M., Wu Y., et al. ssODN-mediated in-frame deletion with CRISPR/Cas9 restores FVIII function in hemophilia A-patient-derived iPSCs and ECs. *Mol. Ther. Nucleic Acids.* 2019; 17: 198-209. doi: 10.1016/j.omtn.2019.05.019

284. Lyu C., Shen J., Wang R., et al. Targeted genome engineering in human induced pluripotent stem cells from patients with hemophilia B using the CRISPR-Cas9 system. *Stem Cell Res. Ther.* 2018; 9(1): 92. doi: 10.1186/s13287-018-0839-8

285. Chen H., Shi M., Gilam A., et al. Hemophilia A ameliorated in mice by CRISPR-based in vivo genome editing of human Factor VIII. *Sci. Rep.* 2019; 9(1): 16838. doi: 10.1038/s41598-019-53198-y

286. Chung J.Y., Ain Q.U., Song Y., et al. Targeted delivery of CRISPR interference system against *Fabp4* to white adipocytes ameliorates obesity, inflammation, hepatic steatosis, and insulin resistance. *Genome Res.* 2019; 29(9): 1442-52. doi: 10.1101/gr.246900.118

287. Maxwell K.G., Augsornworawat P., Velazco-Cruz L., et al. Gene-edited human stem cell-derived β cells from a patient with monogenic diabetes reverse preexisting diabetes in mice. *Sci. Transl. Med.* 2020; 12(540): eaax9106. doi: 10.1126/scitranslmed.aax9106

288. Cho E.Y., Ryu J.Y., Lee H.A.R., et al. Lecithin nano-liposomal particle as a CRISPR/Cas9 complex delivery system for treating type 2 diabetes. *J. Nanobiotechnology.* 2019; 17(1): 19. doi: 10.1186/s12951-019-0452-8

289. Grotz A.K., Abaitua F., Navarro-Guerrero E., et al. A CRISPR/Cas9 genome editing pipeline in the EndoC-βH1 cell line to study genes implicated in beta cell function. *Wellcome Open Res.* 2020; 4: 150. doi: 10.12688/wellcomeopenres.15447.2

290. Olivaes J., Bonamino M.H., Markoski M.M. CRISPR/Cas 9 system for the treatment of dilated cardiomyopathy: A hypothesis related to function of a MAP kinase. *Med. Hypotheses.* 2019; 128: 91-3. doi: 10.1016/j.mehy.2019.05.013

291. Zhao H., Li Y., He L., et al. In vivo AAV-CRISPR/Cas9-mediated gene editing ameliorates atherosclerosis in familial hypercholesterolemia. *Circulation*. 2020; 141(1): 67-79. doi: 10.1161/CIRCULATIONAHA.119.042476

292. Caron J., Pène V., Tolosa L., et al. Low-density lipoprotein receptor-deficient hepatocytes differentiated from induced pluripotent stem cells allow familial hypercholesterolemia modeling, CRISPR/Cas-mediated genetic correction, and productive hepatitis C virus infection. *Stem Cell Res. Ther.* 2019; 10(1): 221. doi: 10.1186/s13287-019-1342-6

293. Wang X., Raghavan A., Chen T., et al. CRISPR-Cas9 targeting of PCSK9 in human hepatocytes *in vivo* — brief report. *Arterioscler. Thromb. Vasc. Biol.* 2016; 36(5): 783-6. doi: 10.1161/ATVBAHA.116.307227

294. Liu X., Zhang Y., Cheng C., et al. CRISPR-Cas9-mediated multiplex gene editing in CAR-T cells. *Cell Res.* 2017; 27(1): 154-7. doi: 10.1038/cr.2016.142

295. Choi B.D., Yu X., Castano A.P., et al. CRISPR-Cas9 disruption of PD-1 enhances activity of universal EGFRvIII CAR T cells in a preclinical model of human glioblastoma. *J. Immunother. Cancer*. 2019; 7(1): 304. doi: 10.1186/s40425-019-0806-7

296. McGowan E., Lin Q., Ma G., et al. PD-1 disrupted CAR-T cells in the treatment of solid tumors: promises and challenges. *Biomed. Pharmacother.* 2020; 121: 109625. doi: 10.1016/j.biopha.2019.109625

297. Guo X., Jiang H., Shi B., et al. Disruption of PD-1 enhanced the anti-tumor activity of chimeric antigen receptor T cells against hepatocellular carcinoma. *Front. Pharmacol.* 2018; 9: 1118. doi: 10.3389/fphar.2018.01118

298. Liu J., Zhou G., Zhang L., et al. Building potent chimeric antigen receptor T cells with CRISPR genome editing. *Front. Immunol.* 2019; 10: 456. doi: 10.3389/fimmu.2019.00456

299. Nakazawa T., Natsume A., Nishimura F., et al. Effect of CRISPR/Cas9-mediated PD-1-disrupted primary human third-generation CAR-T cells targeting EGFRvIII on in vitro human glioblastoma cell growth. *Cells*. 2020; 9(4): 998. doi: 10.3390/cells9040998

300. Ren J., Zhang X., Liu X., et al. A versatile system for rapid multiplex genome-edited CAR T cell generation. *Oncotarget*. 2017; 8(10): 17002-11. doi: 10.18632/oncotarget.15218

301. Ren J., Liu X., Fang C., et al. Multiplex genome editing to generate universal CAR T cells resistant to PD1 inhibition. *Clin. Cancer Res.* 2017; 23(9): 2255-66. doi: 10.1158/1078-0432.CCR-16-1300

302. Zhang Y., Zhang X., Cheng C., et al. CRISPR-Cas9 mediated LAG-3 disruption in CAR-T cells. *Front. Med.* 2017; 11(4): 554-62. doi: 10.1007/s11684-017-0543-6

303. Blaeschke F., Willier S., Stenger D., et al. Leukemia-induced dysfunctional TIM-3^{+}CD4^{+} bone marrow T cells increase risk of relapse in pediatric B-precursor ALL patients. *Leukemia*. 2020. doi: 10.1038/s41375-020-0793-1

304. Stadtmauer E.A., Fraietta J.A., Davis M.M., et al. CRISPR-engineered T cells in patients with refractory cancer. *Science*. 2020; 367(6481): eaba7365. doi: 10.1126/science.aba7365

305. Eyquem J., Mansilla-Soto J., Giavridis T., et al. Targeting a CAR to the TRAC locus with CRISPR/Cas9 enhances tumour rejection. *Nature*. 2017; 543(7643): 113-7. doi: 10.1038/nature21405

306. Dai X., Park J.J., Du Y., et al. One-step generation of modular CAR-T cells with AAV-Cpf1. *Nat. Methods.* 2019; 16(3): 247-54. doi: 10.1038/s41592-019-0329-7

307. MacLeod D.T., Antony J., Martin A.J., et al. Integration of a CD19 CAR into the TCR alpha chain locus streamlines production of allogeneic gene-edited CAR T Cells. *Mol. Ther.* 2017; 25(4): 949-61. doi: 10.1016/j.ymthe.2017.02.005

308. Liu X., Zhao Y. CRISPR/Cas9 genome editing: Fueling the revolution in cancer immunotherapy. *Curr. Res. Transl. Med.* 2018; 66(2): 39-42. doi: 10.1016/j.retram.2018.04.003

309. Roth T.L., Puig-Saus C., Yu R., et al. Reprogramming human T cell function and specificity with non-viral genome targeting. *Nature*. 2018; 559(7714): 405-9. doi: 10.1038/s41586-018-0326-5

310. Song X., Liu C., Wang N., et al. Delivery of CRISPR/Cas systems for cancer gene therapy and immunotherapy. *Adv. Drug. Deliv. Rev.* 2020; S0169-409X(20)30030-2. doi: 10.1016/j.addr.2020.04.010

311. Wang H.X., Li M., Lee C.M., et al. CRISPR/Cas9-based genome editing for disease modeling and therapy: challenges and opportunities for nonviral delivery. *Chem. Rev*. 2017; 117(15): 9874-906. doi: 10.1021/acs.chemrev.6b00799

312. Kotterman M.A., Chalberg T.W., Schaffer D.V. Viral vectors for gene therapy: translational and clinical outlook. *Annu. Rev. Biomed. Eng.* 2015; 17: 63-89. doi: 10.1146/annurev-bioeng-071813-104938

313. Xu C.L., Ruan M.Z.C., Mahajan V.B., et al. Viral delivery systems for CRISPR. *Viruses*. 2019; 11(1): 28. doi: 10.3390/v11010028

314. Wan T., Niu D., Wu C., et al. Material solutions for delivery of CRISPR/Cas-based genome editing tools: current status and future outlook. *Mater. Today.* 2019; 26: 40-66. doi: 10.1016/j.mattod.2018.12.003

315. Pardee K., Green A.A., Takahashi M.K., et al. Rapid, low-cost detection of Zika virus using programmable biomolecular components. *Cell*. 2016; 165: 1255-1266. doi: 10.1016/j.cell.2016.04.059.

316. Müller V., Rajer F., Frykholm K., et al. Direct identification of antibiotic resistance genes on single plasmid molecules using CRISPR/Cas9 in combination with optical DNA mapping. *Sci. Rep.* 2016; 6: 37938. doi: 10.1038/srep37938

317. Guk K., Keem J.O., Hwang S.G., et al. A facile, rapid and sensitive detection of MRSA using a CRISPR-mediated DNA FISH method, antibody-like dCas9/sgRNA complex. *Biosens Bioelectron.* 2017; 95: 67-71. doi: 10.1016/j.bios.2017.04.016

318. Quan J., Langelier C., Kuchta A., et al. FLASH: a next-generation CRISPR diagnostic for multiplexed detection of antimicrobial resistance sequences. *Nucleic Acids Res.* 2019; 47(14): e83. doi: 10.1093/nar/gkz418

319. Chen J.S., Ma E., Harrington L.B., et al. CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science*. 2018; 360(6387): 436-9. doi: 10.1126/science.aar6245

320. Gootenberg J.S., Abudayyeh O.O., Kellner M.J., et al. Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6. *Science*. 2018; 360(6387): 439-44. doi: 10.1126/science.aaq0179

321. Myhrvold C., Freije C.A., Gootenberg J.S., et al. Field-deployable viral diagnostics using CRISPR-Cas13. *Science*. 2018; 360(6387): 444-8. doi: 10.1126/science.aas8836

322. Doench J.G., Fusi N., Sullender M., et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* 2016; 34(2): 184-91. doi: 10.1038/nbt.3437

323. Sanson K.R., Hanna R.E., Hegde M., et al. Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nat. Commun.* 2018; 9(1): 5416. doi: 10.1038/s41467-018-07901-8

324. Kim H.K., Min S., Song M., et al. Deep learning improves prediction of CRISPR-Cpf1 guide RNA activity. *Nat. Biotechnol.* 2018; 36(3): 239-41. doi: 10.1038/nbt.4061

325. Sanson K.R., DeWeirdt P.C., Sangree A.K., et al. Optimization of AsCas12a for combinatorial genetic screens in human cells. *bioRxiv*. 2020. doi.org: 10.1101/747170

326. Montague T.G., Cruz J.M., Gagnon J.A., et al. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res.* 2014; 42(Web Server issue): W401-7. doi: 10.1093/nar/gku410

327. Labun K., Montague T.G., Gagnon J.A., et al. CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. *Nucleic Acids Res.* 2016; 44(W1): W272-6. doi: 10.1093/nar/gkw398

328. Labun K., Montague T.G., Krause M., et al. CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res.* 2019; 47(W1): W171-4. doi: 10.1093/nar/gkz365

329. Haeussler M., Schönig K., Eckert H., et al. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol.* 2016; 17(1): 148. doi: 10.1186/s13059-016-1012-2

330. Heigwer F., Kerr G., Boutros M. E-CRISP: fast CRISPR target site identification. *Nat. Methods.* 2014; 11(2): 122-3. doi: 10.1038/nmeth.2812

331. Bae S., Park J., Kim J.S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics*. 2014; 30(10): 1473-5. doi: 10.1093/bioinformatics/btu048

332. Cradick T.J., Qiu P., Lee C.M., et al. COSMID: a web-based tool for identifying and validating CRISPR/Cas off-target sites. *Mol. Ther. Nucleic Acids.* 2014; 3(12): e214. doi: 10.1038/mtna.2014.64

333. Zhao C., Zheng X., Qu W., et al. CRISPR-offinder: a CRISPR guide RNA design and off-target searching tool for user-defined protospacer adjacent motif. *Int. J. Biol. Sci.* 2017; 13(12): 1470-8. doi: 10.7150/ijbs.21312

334. Kurtz S., Schleiermacher C. REPuter: fast computation of maximal repeats in complete genomes. *Bioinformatics.* 1999; 15(5): 426-7. doi: 10.1093/bioinformatics/15.5.426

335. Abouelhoda M.I., Kurtz S., Ohlebusch E. Replacing suffix trees with enhanced suffix arrays. *J. Discrete Algorithms.* 2004; 2: 53-86. doi: 10.1016/S1570-8667(03)00065-0

336. Biswas A., Fineran P.C., Brown C.M. Accurate computational prediction of the transcribed strand of CRISPR non-coding RNAs. *Bioinformatics*. 2014; 30(13): 1805-13. doi: 10.1093/bioinformatics/btu114

337. Hyatt D., Chen G.L., Locascio P.F., et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics.* 2010; 11: 119. doi: 10.1186/1471-2105-11-119

338. Abby S.S., Néron B., Ménager H., et al. MacSyFinder: a program to mine genomes for molecular systems with an application to CRISPR-Cas systems. *PLoS One.* 2014; 9(10): e110726. doi: 10.1371/journal.pone.0110726

339. Biswas A., Gagnon J.N., Brouns S.J., et al. CRISPRTarget: bioinformatic prediction and analysis of crRNA targets. *RNA Biol.* 2013; 10(5): 817-27. doi: 10.4161/rna.24046

340. Alkhnbashi O.S., Costa F., Shah S.A., et al. CRISPRstrand: predicting repeat orientations to determine the crRNA-encoding strand at CRISPR loci. *Bioinformatics*. 2014; 30(17): i489-96. doi: 10.1093/bioinformatics/btu459

341. Lange S.J., Alkhnbashi O.S., Rose D., et al. CRISPRmap: an automated classification of repeat conservation in prokaryotic adaptive immune systems. *Nucleic Acids Res.* 2013; 41(17): 8034-44. doi: 10.1093/nar/gkt606

342. Raden M., Ali S.M., Alkhnbashi O.S., et al. Freiburg RNA tools: a central online resource for RNA-focused research and teaching. *Nucleic Acids Res.* 2018; 46(W1): W25-9. doi: 10.1093/nar/gky329

343. Nethery M.A., Barrangou R. CRISPR Visualizer: rapid identification and visualization of CRISPR loci via an automated high-throughput processing pipeline. *RNA Biol.* 2019; 16(4): 577-84. doi: 10.1080/15476286.2018.1493332

344. Dion M.B., Labrie S.J., Shah S.A., et al. CRISPRStudio: a user-friendly software for rapid CRISPR array visualization. *Viruses*. 2018; 10(11): 602. doi: 10.3390/v10110602

345. Clement K., Rees H., Canver M.C., et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* 2019; 37(3): 224-6. doi: 10.1038/s41587-019-0032-3

346. Park J., Lim K., Kim J.S., et al. Cas-analyzer: an online tool for assessing genome editing results using NGS data. *Bioinformatics*. 2017; 33(2): 286-8. doi: 10.1093/bioinformatics/btw561

347. Güell M., Yang L., Church G.M. Genome editing assessment using CRISPR Genome Analyzer (CRISPR-GA). *Bioinformatics*. 2014; 30(20): 2968-70. doi: 10.1093/bioinformatics/btu427

348. Brinkman E.K., Chen T., Amendola M., et al. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* 2014; 42(22): e168. doi: 10.1093/nar/gku936

349. Brinkman E.K., Chen T., de Haas M., et al. Kinetics and fidelity of the repair of Cas9-induced double-strand DNA breaks. *Mol. Cell.* 2018; 70(5): 801-13.e6. doi: 10.1016/j.molcel.2018.04.016

CHAPTER 4
Methods of detecting nonspecific activity of genome editing systems

All researchers who use some genome editing system (programmable or engineered nucleases) in their work face the same problems regardless of the target organism: confirming the desired mutation in the target sequence and identifying unintended, yet predictable, mutations outside the target site (off-target sites).

Currently, all programmable nucleases those are used for targeted genome editing/alteration (ZFN, TALEN, CRISPR/Cas) generate mutations outside the target locus at sites that may differ by several nucleotides from the target sequence. The formation of double-strand breaks at off-target genomic loci can lead to an insertion or deletion and to a translocation, which should be carefully monitored [1]. Understanding and avoiding the effects of non-specific editing are important for both fundamental and applied purposes.

Specific methods have been developed to identify off-target mutations; most of them incorporate a specific variant of screening to detect mutations either in predicted regions or anywhere in the genome.

4.1. Biased methods (for off-target mutations' confirmation at the predicted sites)

One of the strategies for detecting off-target sites of the CRISPR/Cas system is based on the search for potential loci binding to the guide RNA, using bioinformatics methods, the i.e., *in silico* prediction. Pre-selected off-target recognition sites of the CRISPR/Cas system are then analyzed by using standard methods such as PCR and heteroduplex analysis or high-throughput sequencing of PCR products. Various research groups have provided extensive data on possible off-target effects induced by different CRISPR/Cas systems. The data obtained were used to develop more accurate algorithms to identify potential off-target sites, such as Cas-OFFinder (www.rgenome.net/cas-offinder) [2], Feng Zhang lab's Target Finder (http://crispr.mit.edu), CasFinder (http://arep.med.har-

vard.edu/CasFinder/), CRISPR Design Tool (http://www.genome-engineering.org) [3], E-CRISP (http://www.e-crisp.org/E-CRISP) [4], Breaking-cas (http://bioinfogp.cnb.csic.es/tools/breakingcas) [5], and others. These algorithms have been used by several research groups to identify potential off-target sites. For example, Y. Chen *et al*. identified candidate genes in human pluripotent stem cell lines that may undergo nonspecific editing. The authors applied the Feng Zhang lab's Target Finder algorithm and analyzed 114 potential off-target sites, none of which were found to contain deletions or insertions [6]. Other authors, using the analysis of data obtained by deep sequencing aimed to identify a limited number of off-target sites, concluded that the editing occurred even in places that are quite different from the intended target site [3, 7, 8].

The ideal PAM (protospacer adjacent motif) sequence recognized by Cas9 is 5'-NGG-3', but this protein can recognize and cleave the sites ending in 5'-NAG-3' or 5'-NGA-3'. Cleavage in close proximity to these sequences is less efficient, but these sequences can act as a PAM-like motif. Notably, DNA cleavage occurs even in the presence of a mismatch of up to 6 nucleotides in the protospacer sequence as well as single nucleotide insertions and/or deletions. Such inconsistencies have a lesser effect on the efficiency of the CRISPR/Cas system when they are located closer to the 5'-end of the protospacer sequence.

Several studies have shown a rather complex picture of Cas9 specificity. The effects of a single mismatch are not always predictable based on their position in the guide RNA only. In addition, the frequency of cleavage is affected not only by the genome but also by some epigenomic factors, making it exceedingly difficult to develop an algorithm that can identify all potential off-target sites. It can be concluded that the above methods require additional research to reliably identify off-targets.

4.2. Unbiased methods (for screening off-target mutations across the whole genome)

4.2.1. Whole genome sequencing

Widespread use of whole-genome sequencing (WGS), especially considering its screening potential, indicates the possibility of WGS application to detect off-target mutations generated during genome editing. However, this method is expensive and can be used for a relatively small number of clones. Sequencing the entire genome allows to detect not only insertions, deletion and single nucleotide substitutions but also the structural variants like inversions, rearrangements (translocations), duplication and large deletions [9]. The limitation of the WGS approach with using a small number of clones for analysis means that most of the off-target rare mutations could be missed [10]. Nevertheless, this approach has been used to identify off-target mutations arising upon CRISPR/Cas9 cleavage for a number of species, including humans (pluripotent stem cells) [11, 12], mouse [13], nematode [14] and plants [15, 16].

4.2.2. Whole exome sequencing

Whole exome sequencing represents a compromise between targeted sequencing and whole genome sequencing to study variants in human genes. Sequencing of all protein-coding regions in a genome allows identifying relevant alternatives of off-target mutations in the exome at a much lower cost than WGS [17]. Depending on the organism, only a small percentage of a genome needs to be covered in this approach but mutations in regulatory or non-coding regions, such as introns, cannot be detected. Thus, exome sequencing is limited by the high false-negative rate, and many off-target mutations could be missed [18, 19].

4.2.3. Chromatin immunoprecipitation sequencing (ChIP-seq)

Chromatin immunoprecipitation and high-throughput sequencing allows detecting all sites of a ribonucleoprotein complex (containing a specific guide RNA and Cas protein) bound to DNA *in vivo*. This indirect method assumes that the presence of a ribonucleoprotein complex at each genomic locus indicates an induced double-strand break (DSB).

It is known that chromatin immunoprecipitation is the most preferred technology for studying protein-DNA binding [20, 21]. The method includes the treatment of living cells with formaldehyde, which causes the formation of covalent crosslinks between DNA and close protein regions, as well as protein-protein crosslinks. After such treatment, chromatin is hydrolyzed, and immunoprecipitation with specific antibodies allows researchers to isolate the DNA regions for binding the intended proteins [22]. The combination of chromatin immunoprecipitation and high throughput sequencing (ChIP-seq) is used to identify the sites where specific proteins are bound to DNA at the whole-genome level (whole genome distribution of binding sites).

Like transcription factors, Cas proteins, including Cas9, can recognize and bind to a DNA region targeted by the guide RNA (**Fig. 4.1**). In this method, the Cas9 protein is replaced by its catalytically inactive form (dCas9), which in combination with a single guide RNA can bind to more than 1,000 DNA loci, although less than 50% of these sites are reliably hydrolyzed by the Cas9 protein [23]. Several studies have confirmed the cleavage at the dCas9 binding sites, while others have found the genome editing at the Cas9 binding sites to be virtually missing. For example, R. Chenchik *et al*. 2014 found that for the 43 loci predicted by the ChIP-seq method for the CRISPR/Cas guide RNA system targeting the Trp53 locus, double-strand breaks were confirmed only at one site in the target and off-target sequences [24]. In another study, K. Kusku *et al*. 2014 mapped the dCas9 binding sites throughout the genome for 12 different guide RNAs and found significant cleavage at 50% of the predicted off-target sites [25]. Thus, this method can provide important information regarding potential binding sites of the CRISPR/Cas9 system, but its results do not always correspond to the actual activity of nucleases at off-target sites.

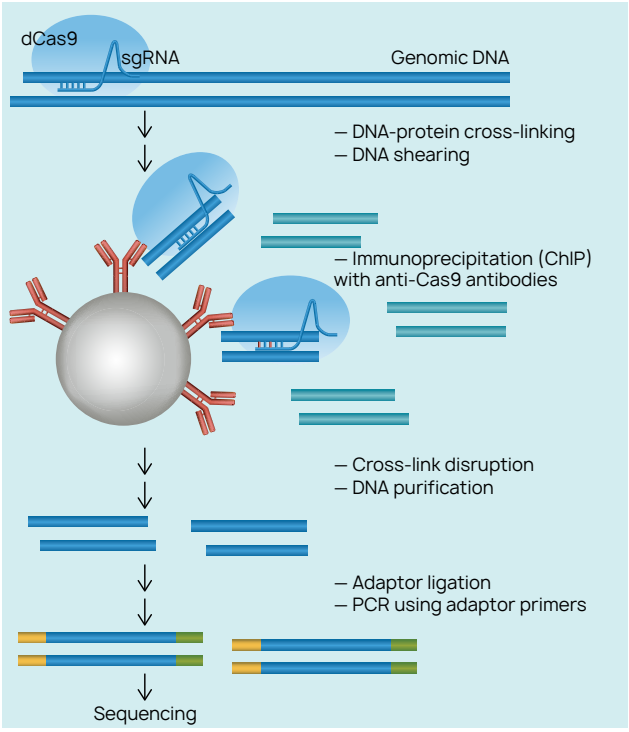


Fig. 4.1. Schematic overview of the ChIP-seq method

4.2.4. Integrase-Defective Lentiviral Vectors (IDLV)

The integrase-defective lentiviral vector (IDLV) technology was one of the first technologies to measure the target and off-target activity of ZFNs across the whole genome [26]. The IDLV type vectors as well as their integrase-competent lentiviral homologues can be delivered into most cells with high efficiency. However, since IDLVs are integrase defective, they remain in the nuclei of the target cells as episomal DNAs. Such episomal vectors can be integrated into sites of double-strand breaks so that they can be used in measuring nonspecific activities of both ZFN and other nucleases including TALE and CRISPR/Cas nucleases [27].

To analyze potential off-target sites, IDLVs are transduced into cells simultaneously with a genome editing system such as ZFN (**Fig. 4.2**). Once a double-strand break has been formed, the IDLV integrates inside it. Then, the genomic DNA is isolated, fragmented to the required size, and the resulting fragments are ligated with adapters. The genomic distribution of the IDLV insertion sites is detected by PCR and subsequent sequencing, and these IDLV insertion sites represent potential off-target loci of the genome [26].

The main advantage of this technology results from the high efficiency of the IDLV insertion into the nuclei of the target cells, including hard-to-transfect primary

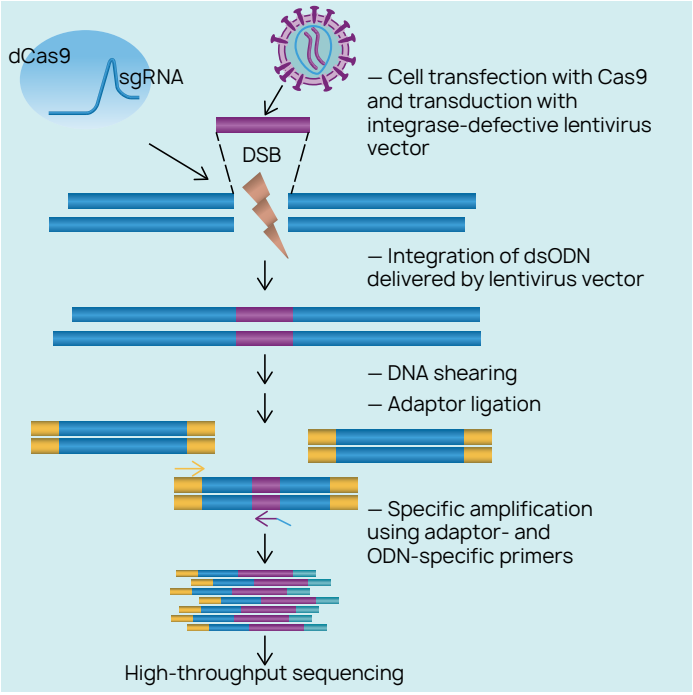


Fig. 4.2. Schematic overview of IDLV method

human cells. However, this method requires the setting of appropriate control samples, since IDLVs can accidentally get inserted into the genome of target cells [28].

4.2.5. Genome-wide unbiased identification of DSBs enabled by sequencing (GUIDE-seq)

Genome-wide unbiased identification of DSBs enabled by sequencing (GUIDE-seq) is based on the insertion of short double-stranded oligodeoxyribonucleotides (dsODNs) (protected by phosphorothioate linkages) at the double-strand break sites (**Fig. 4.3**) [30]. These dsODNs serve as a label for further amplification of genomic sites containing double-strand breaks, which are then identified by high-throughput sequencing. Detection of double-strand breaks is performed by mapping the resulting reads to a reference genome. GUIDE-seq is a quite sensitive method, and it is capable to detect off-target sites occurring at a frequency of 0.1% per cell population.

Inititally this method was used to identify off-target sites of ten different guide RNAs corresponding to different genes of two types of cells. The authors were able to find off-target loci containing up to six mismatches between the expected binding sites of the guide RNA and the target sequence (i.e. in the protospacer sequence) [29]. The GUIDE-seq method demonstrated that cleavage of some off-target sites occurred much more often than the formation of double-strand breaks at the target loci.

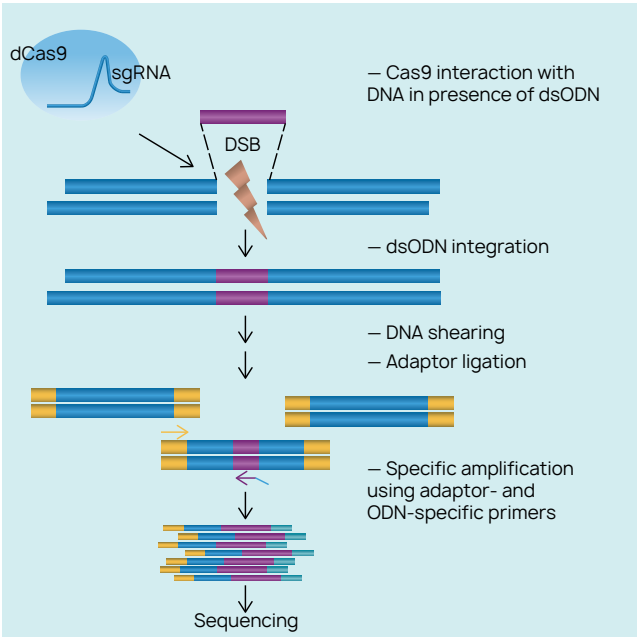


Fig. 4.3. Schematic overview of GUIDE-seq method

The principal advantage of this technology is the accuracy with which dsODN integration can identify off-target sites, as well as the direct correlation between the number of reads at a particular site and the frequency of double-strand breaks induced by the CRISPR/Cas system. The main limitation of this method is associated with the low efficiency of transfection of the target cells [28].

4.2.6. Direct *in situ* breaks labeling, enrichment on streptavidin and next-generation sequencing (BLESS)

The technology of direct *in situ* breaks labeling, enrichment on streptavidin and next-generation sequencing (BLESS) involves the mapping of double-strand breaks across the whole genome by ligation of a biotinylated linker to the free ends of these breaks and subsequent hybridization of biotinylated fragments with streptavidin (**Fig. 4.4**). The labeled fragments are then ligated to a barcoding linker; followed by PCR (with the primers specific to the labels used and sequencing) [30].

The advantage of this method over the approaches based on the labeling of double-stranded breaks is that the breaks themselves are labeled more easily than the proteins associated with them. Several research groups those used BLESS to detect breaks made by different Cas9 variants in mice and human cells have observed a low frequency of off-target activity [31, 32]. Although this method allows mapping breaks across the whole genome, it can identify only the double-strand breaks which were present at the time of their labeling, but not the so-called early breaks that have already undergone reparations [33].

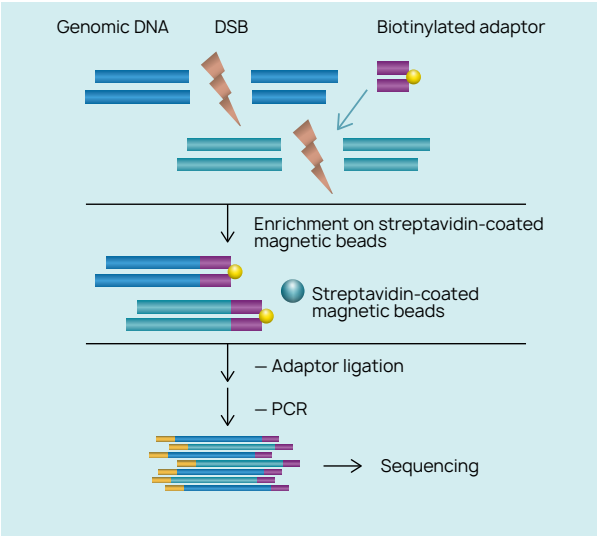


Fig. 4.4. Schematic overview of BLESS method

4.2.7. Linear amplification mediated high-throughput genome-wide translocation-sequencing (LAM HTGTS)

The technology of linear amplification-mediated high-throughput genome-wide translocation-sequencing (LAM HTGTS) was developed to track genomic translocations arising from the repair of double-strand breaks induced by programmable nucleases (TALEN and Cas9) [34]. Based on this method, a programmed nuclease introduced into the cell cleaves the bait sequence (DSB primers) and the resulting break is repaired by fusion with another double-stranded break (**Fig. 4.5**). As a result of this process, chromosomal translocations can occur if the breaks are located in different chromosomes or at different loci of the same chromosome. The genomic DNA is then isolated and fragmented by ultrasound. Since the bait sequence is known, the linear PCR with a primer-specific biotinylated oligonucleotide can amplify a translocation region including an unknown double-stranded break. The biotinylated single-stranded DNA is purified by streptavidin-coated magnetic beads. Adapters are ligated to the end of the purified biotinylated DNA, followed by PCR. The amplified fragments obtained are subjected to sample preparation for subsequent high-throughput sequencing. The repaired double-strand breaks those have not undergone translocation carry a restriction endonuclease recognition site, which can be used for selective cleavage, and they will not be amplified or sequenced.

Standard LAM-HTGTS technology fails to identify small insertions, deletions, or single nucleotide substitutions, but it can be modified to fullfill these needs. In this case, a greater sequencing depth is required to compensate for the higher number of reads of non-translocated mutations and the bait sequence itself. LAM-HTGTS technology is a qualitative method to screen cells for large genomic rearrangements

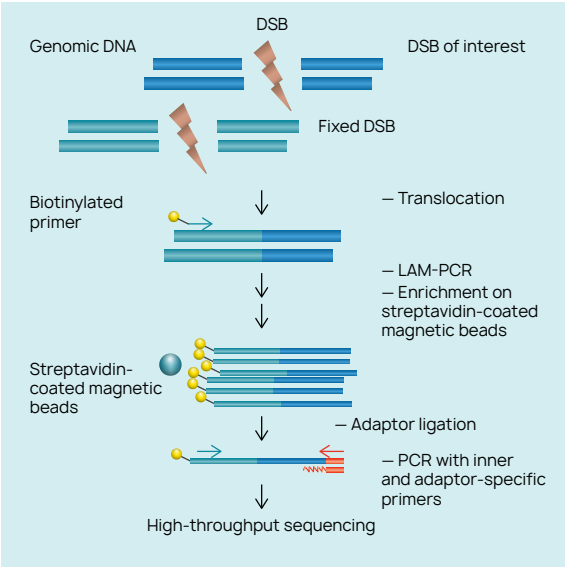


Fig. 4.5. Schematic overview of the LAM-HTGTS method

caused by the programmed nucleases in both target and off-target sequences. However, it should be noted that upon formation of the double-strand breaks translocations occur at a rather low frequency constituting one translocation per 300 cells according to some estimates which indicates the limitations of this method for reliable detection of all off-target loci [35].

4.2.8. *In vitro* Cas9-digested whole genome sequencing (Digenome-seq)

Cas9 protein can be used not only as a genome-editing tool but also as a nuclease for *in vitro* profiling off-target effects of selected guide RNAs. This forms the basis for *in vitro* Cas9-digested whole genome sequencing (Digenome-seq), where the extracellular genomic DNA is cleaved by the Cas9 protein *in vitro*, and the resulting fragments are sequenced by next-generation sequencing (NGS) (**Fig. 4.6**). The genomic DNA is isolated from two types of transfected cells — both with and without introduced nucleases. The resulting DNA is modified using CRISPR/Cas9 *in vitro*. For unmodified genes, double-strand breaks will be inserted at precise (specific) positions, thus representing forward alignment profiles after sequencing [36]. The sequences of genes carrying mutations will differ from those of the unmodified genes which cannot be recognized by Cas9 *in vitro*, resulting in a stepwise alignment upon sequencing. Thus, the Digenome-seq technology makes it to create a list of potential off-target CRISPR/Cas nuclease recognition sites based on the sequence alignment profiles.

Digenome-seq technology can also be used in a multiplexing format, which makes it possible to analyze up to 10 guide RNAs at a time. The advantages of this

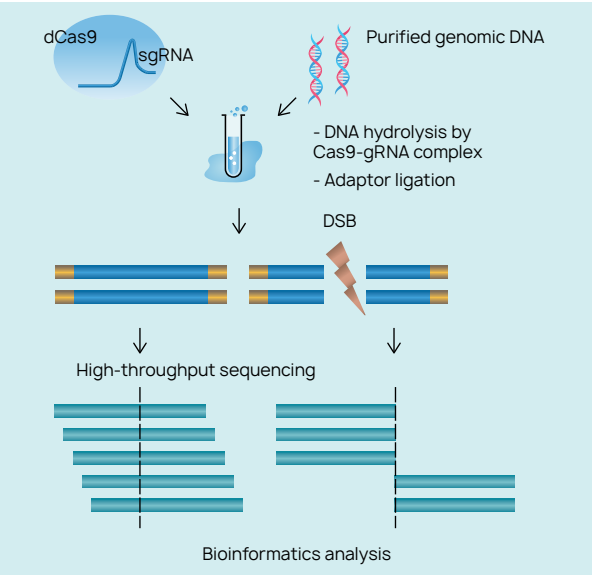


Fig. 4.6. Schematic overview of Digenome-seq method

method are the lack of amplification requirement and the fact that the breaks obtained *in vitro* are not subject to reparation. In addition, this technology does not detect incidentally arose double-strand DNA breaks in a cell [37]. However, the artificial environment of extracellular genomic DNA, in turn, gives rise to a potential drawback of Digenome-seq technology: the differences between *in vitro* and *in vivo* activity and Cas9 specificity can lead to false positive or false negative results [38].

To check whether the structure of chromatin in eukaryotic cells affects the target and the activity of CRISPR/Cas9, the authors of the Digenome-seq technology first identified a series of identical endogenous DNA sequences present in regions with both high and low degree of chromatin compaction, and then measured mutation frequencies for these sites in human cells, using Cas9 in combination with completely complementary (precise) and substituted guide RNAs [39]. Unlike precise guide RNAs, the RNAs containing substitutions were highly sensitive to chromatin, which suggested that the chromatin prevents cleavage of DNA outside the target. Then, using the Digenome-seq technique performed in parallel on cell-free chromatin DNA and histone-free genomic DNA, it was shown that only a small part of double-strand breaks found in genomic DNA could be identified in chromatin DNA. Thus, the chromatin can inhibit the off-target effects of Cas9 across the genome in the cells facilitating its specificity.

4.2.9. Selective enrichment and identification of tagged genomic DNA ends by sequencing (SITE-seq)

The technology of selective enrichment and identification of tagged genomic DNA ends by sequencing (SITE-seq), like BLESS, is based on tagging double-strand

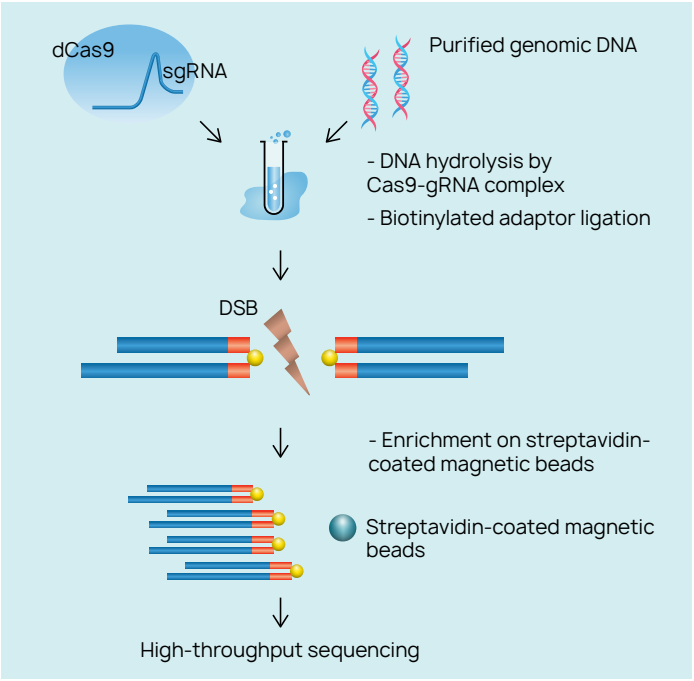


Fig. 4.7. Schematic overview of SITE-seq method

breaks. The isolated genomic DNA undergoes *in vitro* modification by CRISPR/Cas9 (**Fig. 4.7**) to introduce of double-strand breaks in both on-target and off-target regions. A biotinylated adapter is ligated to the free ends of the breaks; DNA fragmentation is followed by ligation of the second adapter and hybridization on streptavidine-coated magnetic beads. Then amplification of enriched fragments followed by high-throughput sequencing is performed [40].

Unlike BLESS technology, SITE-seq does not depend on endogenous reparation. Thus, using this technology, it is possible to identify a greater number of off-target sites, which, in turn, are detected in accordance with the alignment profiles (patterns), thanks to which the required coverage of the analyzed sequences can be significantly reduced [37].

4.2.10. Circularization for in vitro reporting of cleavage effects by sequencing (CIRCLE-seq)

Circularization for *in vitro* reporting of cleavage effects by sequencing (CIRCLE-seq) is more sensitive than the previously described Digenome-seq [41]. For the CIRCLE-seq, the isolated fragmented genomic DNA is locked in a covalently closed circle (i.e., it is circularized) and cleaved by CRISPR/Cas9 *in vitro* (**Fig. 4.8**). When the Cas9 nuclease interacts with both target and off-target loci, circular DNAs are converted into linear structures with free ends, to which adapters are ligated, followed by PCR with subsequent sequencing.

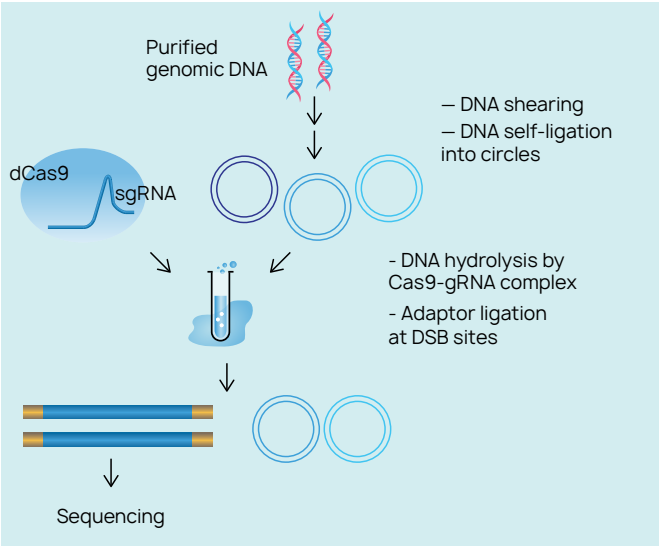


Fig. 4.8. Schematic overview of the CIRCLE-seq method

It should be noted that the authors who developed this technology used the strategies independent of restriction endonucleases for the formation and subsequent enzymatic selection of one of the two types of covalently closed DNA structures, namely, the attachment of a *hairpin* to the ends of DNA or circularization of linear DNA fragments. Comparison of these two approaches showed that circularization is dramatically more efficient in the enrichment of genomic DNA fragments cleaved by the Cas9 nuclease. Notably, nearly all the identified sites revealed by using fragments of linear DNA with the ends closed in hairpins were also detected by using circularized DNA, and the numbers of reads in both cases were strongly correlated. This suggested that circularization did not shift the range or frequency of the identified off-target sites.

CIRCLE-seq technology is characterized by a fairly low percentage of background reads (~1.7%) in the total number of resulting reads, which allows researchers to reliably identify off-target sites using this technology. The results obtained with this method can be analyzed without a reference sequence.

Currently the CIRCLE-seq method is a highly sensitive and the most efficient, in terms of sequencing, *in vitro* approach to the genome-wide identification of off-target CRISPR/Cas9 nuclease cleavage sites [41]. Compared to CIRCLE-seq, reads of SITE-seq and Digenome-seq cover only half of the cleavage sites. Compared to Digenome-seq, CIRCLE-seq virtually eliminates the high background of random reads. In addition, the CIRCLE-seq is more sensitive than cell-based methods for off-target sites detection. In most cases, CIRCLE-seq can identify all off-target sites in human genomic DNA found by GUIDE-seq, which is one of the most sensitive cell-based methods. Besides, CIRCLE-seq also identified new reliable off-target sites which occur in human cells, demonstrating that it can detect new off-target loci lying outside the GUIDE-seq detection limits [42].

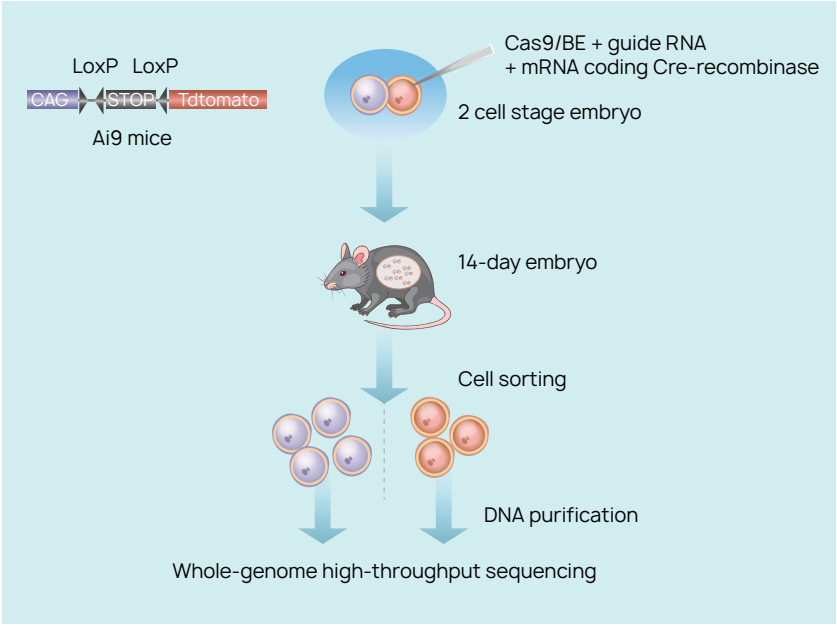


Fig. 4.9. Schematic overview of GOTI method

It should be noted that each sample for CIRCLE-seq requires a relatively large amount of genomic DNA for 'circularization' (~25 μg), which may represent a limiting factor, depending on the availability of the cells studied.

4.2.11. Genome-wide off-target analysis by two-cell embryo injection (GOTI)

The authors of Genome-wide off-target analysis by two-cell embryo injection (GOTI) edited one blastomere of a two-cell embryo using CRISPR/Cas9 and labeled it with a red fluorescent protein (**Fig. 4.9**). On the 14th day of development, the whole mouse embryo was divided into individual cells. Sorting of the edited cells and wild-type cells was performed using flow cytometry. The resulting groups of cells were sequenced by NGS followed by bioinformatic comparative analysis [43].

It should be noted that this *in vivo* analysis can be used to avoid *«noise effects»* caused by *in vitro* cell expansion. Moreover, since both experimental and control groups belong to the same fertilized egg, this can exclude the influence of the genetic background.

High editing efficiency, ease of use and low cost are the main contributors to the popularity of CRISPR/Cas systems in both academic and translational applications. However, it is important to note that the major limiting factor for the clinical use of the programmed nucleases (CRISPR/Cas, TALEN, ZFN) is the presence of off-target effects. The identification of both target and off-target cleavage sites is critical not only for understanding the potential side effects of genome-editing technologies but also for development of the new systems with greater specificity.

Table 4.1. Comparison of methods for defining genome-wide CRISPR–Cas9 specificity (from Lazzarotto C.R., Nguyen N.T., Tang X., et al. Defining CRISPR-Cas9 genome-wide nuclease activities with CIRCLE-seq. Nat. Protoc. 2018; 13(11): 2615-42)

Method	Description	Advantages	Limitations
IDLV	Cell-based method where integrase-defective lentiviral vectors are integrated with a selective marker into sites of nuclease-induced DSBs. Vector integration sites are enriched by linear amplification, followed by high-throughput sequencing	Certain cell types may be more amenable to infection with IDLV than transfection with a double-stranded oligodeoxynucleotide (dsODN) tag	Relatively insensitive due to the low integration efficiency and the requirement for enrichment to overcome it. High background level as IDLVs still retain some capability to randomly integrate into the cellular genome in the absence of nuclease-induced double-strand breaks. An IDLV integration can occur at some distance from the nuclease-induced break so it may be more challenging to map the relevant sites
GUIDE-seq	Based on efficient integration of double-stranded oligodeoxynucleotide (dsODN) tags into DSBs by NHEJ in living cells, followed by tag-specific amplification and high-throughput sequencing	High efficiency of label integration (dsODN) into double-strand breaks improves sensitivity. Quantitative correlation between the number of GUIDE-seq reads and the mutation frequencies in living cells	Requires efficient cellular transfection of the dsODN tag, which can be challenging in sensitive cell types or *in vivo* settings
HTGTS	Detects off-target nuclease-induced double-strand breaks by observation of translocation junctions between two nuclease-induced double-stranded breaks	Can be applied to detect off-target nuclease activity when nucleases are delivered *in vivo*	Nuclease-induced translocations are rare. Translocations occur more frequently with sites in the same chromosome or in close nuclear proximity
BLESS	Based on *in situ* ligation of adapters to the ends of nuclease-induced double-strand breaks in fixed cells	Does not require delivery and incorporation of exogenous DNA for detection	Lack of information about nuclease-induced double-stranded breaks that were previously repaired by the cell repair system
DIGENOME-seq	*In vitro* method based on the detection of Cas9-digested genomic DNA by whole genome sequencing	Does not require PCR; has also tested with base editors	Does not enrich for nuclease digested sequences and requires a large number of sequencing reads (~ 400 million). High-level background. Lacks information about how cellular factors affect off-target nuclease activity

Table 4.1 to be continued on page 137.

SITE-seq	*In vitro* based on Cas9-cleavage of high-molecular-weight DNA, followed by enzymatic fragmentation, biotinylated adapter ligation, enrichment and sequencing	Enriches for nuclease cleaved fragments; reduces the number of sequencing reads required	Reads contain only one half of the cleaved sites. Lacks information about how cellular factors affect off- target nuclease activity
CIRCLE-seq	*In vitro* method, in which genomic DNA is randomly fragmented, followed by circularization and generation of covalently closed double-stranded DNA molecules. Circular dsDNAs are cleaved by Cas9 at on-target and off-target sites, allowing the selective sequencing of nuclease-induced DSBs	High enrichment, so fewer reads are required (3-5 million reads); reads contain both halves of the cleavage sites	Lacks information about how cellular factors affect off-target nuclease activity; requires a large amount of genomic DNA

The methods most commonly used to assess the off-target activity of nucleases are presented in **Table 4.1**. All these methods can be divided into two major categories: cell-based and *in vitro* approaches. In general, cell-based methods can directly detect off-target sites those are cleaved in a particular cell type; however, such methods are characterized by increased requirements to the cells studied namely, their ability to cultivate and transfect. In contrast, *in vitro* methods can be more sensitive and more scalable than cell-based methods. Both types of methods ultimately require further validation *in vivo* to confirm whether the off-target sites identified by these methods lead to *reliable* (true) mutagenesis in the cells or tissues modified by genome-editing nucleases [42].

All the methods reviewed in this chapter have their own advantages and disadvantages, and the best method should be chosen carefully for a particular genome-editing experiment. *In vitro* methods such as CIRCLE-seq, Digenome-seq, and SITE-seq are the most complete. However, off-target mutations detected by them should be confirmed in subsequent cell experiments. The methods like GUIDE-seq directly quantify off-target mutations in cells but have limitations in consistent detection of the sites with a low mutation rate (< 0.1%). Ligation-based methods can help understand the kinetics of on-target and off-target genome editing.

Algorithms for guide RNAs engineering will be improved with increasing knowledge of the mechanisms of on-target and off-target activity, while combining the nucleases with a high reproduction accuracy and optimally designed guide RNAs will further enhance the precision of the CRISPR/Cas9 system.

References

1. Sander J.D., Joung J.K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* 2014, 32(4): 347-55.
2. Bae S., Park J., Kim J.S. Cas-offinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics.* 2014, 30(10): 1473-5.
3. Hsu P.D., Scott D.A., Weinstein J.A., et al. DNA targeting specificity of RNA-guided CAS9 nucleases. *Nat. Biotechnol.* 2013, 31(9): 827-32.
4. Heigwer F., Kerr G., Boutros M. E-CRISP: fast CRISPR target site identification. *Nat. Methods*. 2014; 11(2): 122-3.
5. Oliveros J.C., Franch M., Tabas-Madrid D., et al. Breaking-Cas-interactive design of guide RNAs for CRISPR-Cas experiments for ensembl genomes. *Nucleic Acids Res.* 2016; 44(W1): W267-71.
6. Chen Y., Cao J., Xiong M., et al. Engineering human stem cell lines with inducible gene knockout using CRISPR/Cas 9. *Cell Stem. Cell.* 2015; 17(2): 233-44.
7. Fu Y., Foden J.A., Khayter C., et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* 2013; 31(9): 822-6.
8. Ran F.A., Hsu P.D., Lin C.Y., et al. Double nicking by RNA-guided CRISPRCas9 for enhanced genome editing specificity. *Cell.* 2013; 154(6): 1380-9.
9. Veres A., Gosis B.S., Ding Q., et al. Low incidence of off-target mutations in individual CRISPR-Cas9 and TALEN targeted human stem cell clones detected by whole-genome sequencing. *Cell Stem. Cell.* 2014; 15(1): 27-30.
10. Wu X., Kriz A.J., Sharp P.A., et al. Target specificity of the CRISPR-Cas9 system. *Quant. Biol.* 2014; 2(2): 59-70.
11. Yang L., Grishin D., Wang G., et al. Targeted and genomewide sequencing reveal single nucleotide variations impacting specificity of Cas9 in human stem cells. *Nat. Commun.* 2014; 5: 5507.
12. Smith C., Gore A., Yan W., et al. Whole-genome sequencing analysis reveals high specificity of CRISPR/Cas9 and TALEN-based genome editing in human iPSCs. *Cell Stem Cell.* 2014; 15(1): 12-3.
13. Iyer V., Shen B., Zhang W., et al. Offtarget mutations are rare in Cas9-modified mice. *Nat. Methods.* 2015; 12(6): 479.
14. Paix A., Wang Y., Smith H.E., et al. Scalable and versatile genome editing using linear DNAs with microhomology to Cas9 sites in Caenorhabditis elegans. *Genetics.* 2014; 198(4): 1347-56.
15. Feng Z., Mao Y., Xu N., et al. Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in Arabidopsis. *Proc. Natl. Acad. Sci. USA.* 2014; 111(12): 4632-7.
16. Zhang H., Zhang J., Wei P., et al. The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation. *Plant Biotechnol. J.* 2014; 12(6): 797-807.
17. Ng S.B., Turner E.H., Robertson P.D., et al. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*. 2009; 461(7261): 272-6.
18. Cho S.W., Kim S., Kim Y., et al. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res.* 2014; 24(1): 132-41.
19. Karakoc E., Alkan C., O'Roak B.J., et al. Detection of structural variants and indels within exome data. *Nat. Methods.* 2012; 9(2): 176-8.
20. Solomon M.J., Larsen P.L., Varshavsky A., et al. Mapping protein DNA interactions in vivo with formaldehyde: evidence that histone H4 is retained on a highly transcribed gene. *Cell*. 1988; 53(6): 937-47.

21. Mundade R., Ozer H.G., Wei H., et al. Role of chip-seq in the discovery of transcription factor binding sites, differential gene regulation mechanism, epigenetic marks and beyond. *Cell Cycle.* 2014; 13(18): 2847-52.
22. Dedon P.C., Soults J.A., Allis C.D., et al. A simplified formaldehyde fi xation and immunoprecipitation technique for studying protein-DNA interactions. *Anal. Biochem.* 1991; 197(1): 83-90.
23. Wu X., Scott D.A., Kriz A.J., et al. Genomewide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat. Biotechnol.* 2014; 32(7): 670-6.
24. Cencic R., Miura H., Malina A., et al. Protospacer adjacent motif (PAM)-distal sequences engage CRISPR Cas9 DNA target cleavage. *PLoS One.* 2014; 9(10): e109213.
25. Kuscu C., Arslan S., Singh R., et al. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat. Biotechnol.* 2014; 32(7): 677-83.
26. Gabriel R., Lombardo A., Arens A., et al. An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat. Biotechnol.* 2011; 29(9): 816.
27. Wang X., Wang Y., Wu X., et al. Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENS using integrase-defective lentiviral vectors. *Nat. Biotechnol.* 2015; 33(2): 175-8.
28. Martin F., Sanchez-Hernandez S., Gutierrez-Guerrero A., et al. Biased and unbiased methods for the detection of off-target cleavage by CRISPR/Cas9: an overview. *Int. J. Mol. Sci.* 2016; 17(9): 1507.
29. Tsai S.Q., Zheng Z., Nguyen N.T., et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol*. 2015; 33(2): 187-97.
30. Crosetto N., Mitra A., Silva M.J., et al. Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. *Nat. Methods.* 2013; 10(4): 361-5.
31. Ran F.A., Cong L., Yan W.X., et al. In vivo genome editing using Staphylococcus aureus Cas9. *Nature.* 2015; 520(7546): 186-91.
32. Slaymaker I.M., Gao L., Zetsche B., et al. Rationally engineered Cas9 nucleases with improved specificity. *Science*. 2016; 351(6268): 84-8.
33. Tsai S.Q., Joung J.K. Defining and improving the genome-wide specificities of CRISPR-Cas9 nucleases. *Nat. Rev. Genet.* 2016; 17(5): 300-12.
34. Hu J., Meyers R.M., Dong J., et al. Detecting DNA double-stranded breaks in mammalian genomes by linear amplification-mediated high-hroughput genome-wide translocation sequencing. *Nat. Protoc.* 2016; 11(5): 853-71.
35. Roukos V., Voss T.C., Schmidt C.K., et al. Spatial dynamics of chromosome translocations in living cells. *Science*. 2013; 341(6146): 660-4.
36. Kim D., Bae S., Park J., et al. Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat. Methods*. 2015; 12(3): 237.
37. Kim D., Kim S., Kim S., et al. Genome-wide target specificities of CRISPR-Cas9 nucleases revealed by multiplex Digenome-seq. *Genome Res.* 2016; 26(3): 406-15.
38. Fu B.X.H., St. Onge R.P., Fire A.Z., et al. Distinct patterns of Cas9 mismatch tolerance in vitro and in vivo. *Nucleic Acids Res.* 2016; 44(11): 5365-77.
39. Kim D., Kim J.S. DIG-seq: a genome-wide CRISPR off-target profiling method using chromatin DNA. *Genome Res.* 2018; 28(12): 1894-900.
40. Cameron P., Fuller C.K., Donohoue P.D., et al. Mapping the genomic landscape of CRISPR-Cas9 cleavage. *Nat. Methods.* 2017; 14(6): 600-6.
41. Tsai S.Q., Nguyen N.T., Malagon-Lopez J., et al. CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR–Cas9 nuclease off-targets. *Nat. Methods.* 2017; 14(6): 607.
42. Lazzarotto C.R., Nguyen N.T., Tang X., et al. Defining CRISPR-Cas9 genome-wide nuclease activities with CIRCLE-seq. *Nat. Protoc.* 2018; 13(11): 2615-42.
43. Zuo E., Sun Y., Wei W., et al. Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science.* 2019; 364(6437): 289-92.

CHAPTER 5
Nosologic and therapeutic aspects of editing animal genome

Before they become widely available, gene therapy techniques, like any other drug development process, have to be thoroughly tested and tried, including preclinical testing on animal models.

Animal models, most valuable for studying the efficacy of gene therapy, are those that most accurately reproduce the physiological and pathological processes, which may occur in the human body.

Most of the animals used for modeling are organisms that were genetically modified to some extent. It is not unusual, when, to simulate a particular disease, different lines of rodents, for example, are crossed during one experiment. In recent years, transgenic animals have been frequently developed for specific research purposes. Laboratory-animal breeding centers offer transgenic, knockout rodent lines with knocked-out genes at the customer's choice, animals with overexpression (knock-in) or with reduced expression (knockdown) of any gene. In addition, when needed, many researchers independently develop animals with gene mutations required for their scientific research.

Given the diversity of diseases that can be treated with gene therapy, a variety of animal species are used to develop models, from fruit flies (*Drosophila melanogaster*) to cattle or horses.

Comparative studies have found that the human and mouse genomes are approximately 70–90% identical, but with a lot of key variations (for example, some mouse and human gene products are almost identical, while others are hardly recognizable as close relatives); the similarity between the human and the dog genomes is 75%, while the pig genome is 71% identical to the human genome. Cats and primates share a 73% homology with the human genome. Among the 768 identified genes of cattle, 638 (83%) can be considered identical to human genes. All rabbit chromosome paints, except for the Y paint, are hybridized to human chromosomes, and all human chromosome paints, except for the Y paint, are hybridized to rabbit chromosomes. The DNA sequence of the chimpanzee is 98.8% identical to that in humans while

the fruit fly genome and the human genome have 60% of the genes in common. We also share around 50% of the genes with the banana [1].

Despite their genomic similarities to humans, most model organisms typically do not suffer from the same genetic diseases as humans. Therefore, scientists have to find ways to alter animal genomes to induce human diseases. Thus, in attempting to construct a genetic animal model for human disease, it is important to know what kind of mutation causes the disease to induce it in the corresponding animal gene.

Transgenic animals are used to simulate various human diseases, such as neurodegenerative disorders, arteriosclerosis, oncological and infectious diseases, and many others.

Scientists use two approaches for developing animal models. The first approach is targeted and disease driven, while the other one is non-targeted and mutation-free. The targeted disease-based approach can employ any of the available techniques, depending on the specific type of mutation involved in the studied disease. The general targeted techniques include transgenesis, single-gene knock-ins, knockdowns or knockouts, conditional gene modifications, and chromosomal rearrangements. The non-targeted, mutation-driven method of generating mutations implies using radiation and chemicals. The large-scale mutation screening is one of the most common techniques associated with this method [2].

Among large-scale mutation screening techniques and methods, exposure to X-rays or chemical N-ethyl-N-nitrosourea (ENU) mutagen is the most efficient method to induce mutations. X-rays often cause large deletion and translocation mutations involving multiple genes. In its turn, the ENU treatment involves mutations within single genes, such as point mutations; therefore, this method is frequently used in model organisms such as zebrafish (*Danio rerio*). These types of models are especially useful in identifying new pathways and genes contributing to disease. Thus, these methods involve screening to assess the relationship between phenotypes of the obtained models and human diseases by using indirect approaches to inducing mutations in genomes of animal models.

The insertion of complete human genes into genomes of the respective animal species offers an attractive combination of advantages for each system, which can be particularly useful for testing human-specific elements of gene function, regulation, or response to potential therapeutic agents for disease or condition. This approach often involves the insertion of extended DNA sequences with promoters, untranslated flanking regions and introns, which most often exceed the throughput of commonly used plasmid and viral systems. Transgenesis can be performed either by inserting the foreign DNA directly into the embryo or by using a retroviral vector to insert a transgene into the DNA of the organism. The first mouse gene transfers were performed in 1980; however, at that time, the transgenesis methods were not optimal. For example, the foreign DNA was incorporated into only a small percentage of embryos and was inconsistently passed on to the next generation. Furthermore, small transgenes were inserted into random sites in the genome and depending on their location, they were not always expressed. Recently, scientists have developed a method to increase the size of the DNA fragments used in transgenesis by cloning

them in bacterial artificial chromosomes (BACs) or yeast artificial chromosomes (YACs). Three different methods can be used to deliver BAC or YAC DNA to the germline of mice: pronuclear injection of purified DNA, lipofection of DNA into embryonic stem cells, or fusion of yeast spheroplasts with embryonic stem cells. Among these three methods, microinjection of isolated BAC or YAC DNA into the zygote is the most common method for obtaining transgenic animals [3].

Methods employing knock-out and knock-in models target a mutation at a specific gene locus. These methods are particularly useful if a single gene is found to be the primary cause of the disease. Knock-out mice carry a gene that has been inactivated to reduce the expression and cause the loss of function; knock-in mice are produced by inserting a transgene into an exact location where it is overexpressed. Over the years, more than 3,000 genes have been knocked out in mice, and most of these genes have been related to diseases [4].

Both knock-out and knock-in animals are developed in the same way: a specific mutation is induced into the endogenous gene, and then it is conveyed to the next generation through breeding. In some cases, this procedure requires embryonic stem cells (**Fig**. **5.1**), and in others, it is easier and more convenient to use the CRISPR/Cas technology of targeted genome editing (**Fig. 5.2**). One of the distinctive characteristics of embryonic stem cells is their ability to introduce alterations into all cell populations when injected into blastocysts; in addition, they can be genetically modified and selected for the desired gene alterations.

Homologous recombination results in mutations. During this process, two strands of DNA are physically rearranged for the exchange of genetic material. Many types of mutations can be induced to a model gene in this way, including null or point mutations as well as complex chromosomal rearrangements such as large deletions, translocations, or inversions. Many knockout and knock-in mice have similar phenotypes to humans and, therefore, are good models for human diseases.

One of the drawbacks of using transgenic, knockout and knock-in mice to study human diseases is that many disorders occur late in life, and when genes are altered to model such diseases, mutations can profoundly affect the development and cause early death. Such effects prevent using animal models in studying adult diseases in humans. Fortunately, the new technology made it possible to generate mutations in specific tissues and at different stages of development, including adulthood. The method involves mice with two different types of genetic alterations: the first one contains a conditional vector, which is like an *on switch* for the mutation, and the other contains specific sites (*loxP*) inserted on either side of a whole gene, or part of a gene, which encodes a certain component of a protein that will be deleted. A conditional vector for the gene is made by inserting the recognition sequences for the bacterial Cre recombinase (*loxP* sites) by using homologous recombination in embryonic stem cells. The vector contains a drug (typically, antibiotic) resistant marker gene, which allows only the targeted embryonic stem cells to survive when exposed to the drug. Thus, the mutant embryonic stem cells can be selected and injected into a host mouse embryo, which is transferred to a surrogate mouse.

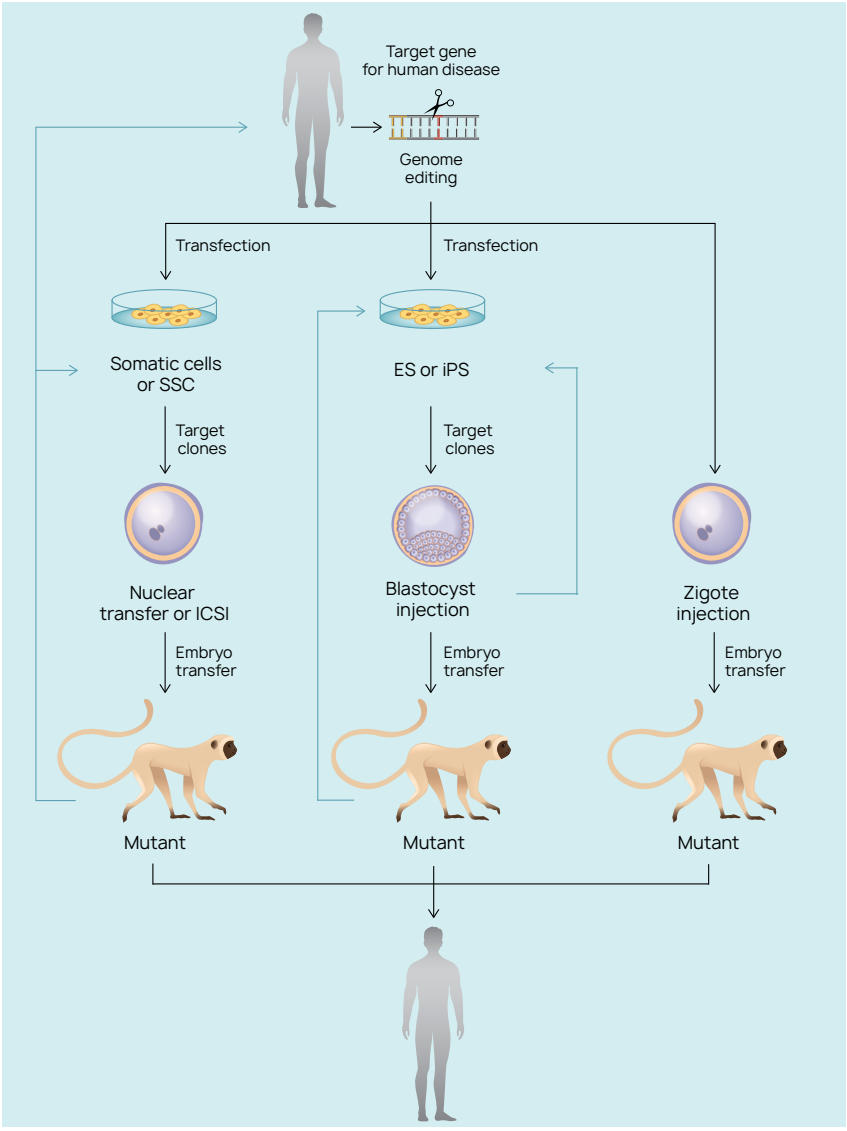


Fig. 5.1. Methods for modifying genes in primates to study human diseases through injection of embryonic or pluripotent stem cells into blastocysts, transfer of nuclei of somatic stem cells, or intracytoplasmic injection of spermatogonial stem cells into the cytoplasm of the zygote

The resulting offspring are chimeras and have multiple populations of genetically distinct cells. The chimeric offspring are then crossed, and the resulting generation of the offspring has a recombinase effector gene. The mice containing the Cre recombinase under the control of tissue-specific or inducible regulatory elements are crossed with mice having the desired *lox*P sites. When Cre is expressed, recombination

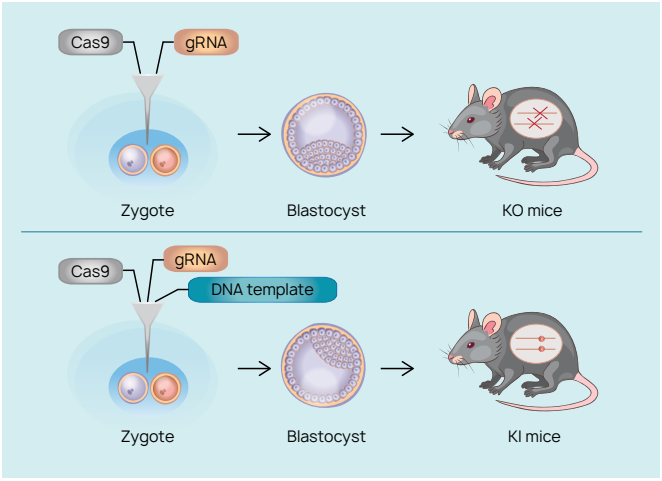


Fig. 5.2. Schematic representation of obtaining knockout and knock-in mice by using CRISPR/Cas technology.
a — injection of guide RNA (gRNA) and Cas 9 will produce indels, which can cause a shift in the translational reading frame and premature termination of protein synthesis (due to the appearance of a stop codon), thereby developing knockout (KO) mice. *b* — addition of a highly homologous DNA template containing a specific mutation will result in knock-in (KI) mice through the process of homology-directed repair. Reagents are injected into the cytoplasm of the zygote. Alternatively, they can be injected into the pronucleus of the zygote, but cytoplasmic injection is simpler and less toxic

occurs at *lox*P sites, which delete the intervening sequences, and the resulting mutation is induced in specific regions at a specific time [5].

The above-mentioned advances in the application of embryonic stem cells and Cre/*loxP* conditional mutations have helped pave the way for constructing models for complex human diseases involving chromosomal rearrangements. Murine models of these disorders can be developed using indirect approaches, such as radiation, but their usefulness is limited, as pathological endpoints are unpredictable and undefined. The application of the Cre/*loxP* recombination system helps overcome these drawbacks by inducing site-specific mutations essential for obtaining accurate models of defects caused by human chromosomal rearrangements. These mutations can include chromosome deletions, duplications, inversions, and translocations as well as chromosome insertions.

With all the diversity of animal species that can serve as point mutation models, some of them are more suitable for developing a model of a specific disease. Fish and non-human primate models deserve special attention. Fish are primarily used as initial models of gene delivery. Nevertheless, it has been found that they can be used as animal models of human diseases such as hepatic diseases or cancer [6]. In addition, the data have shown that zebrafish models can be used to simulate vascular [7, 8] and cardiac [9, 10] development, cardiac regeneration [11] and congenital cardiomyopathy [12]. On the other hand, modeling of diseases in nonhuman primates, when required, usually

takes place at the final stages of research, after positive results have been obtained in other animal models. It is explained by the high cost of these models, as manipulating genes in monkeys is far more difficult than in other animals.

5.1. Fish

The zebrafish is a good model for a better understanding of human congenital diseases such as achromatopsia, campomelic dysplasia or defects in neural crest development. Congenital achromatopsia is caused by mutations in one of a few genes, for example, in the cone transducin gene (*GNAT2*). The no-optokinetic response f^{w21} (*nof*) zebrafish mutants were used to demonstrate that this disorder can be overcome by directing targeted expression of the wild-type protein. Plasmid DNAs were used as vectors in this study [13].

Another example is Waardenburg syndrome (WS) characterized by telecanthus heterochromia iridis, congenital hearing loss, and lack of pigmentation in separate hair strands. These disorders are caused by mutations in *PAX3*, *MITF*, *WS2B*, *WS2C*, *SNAI2*, *SOX10*, *EDNRB*, and *EDN3* genes. However, a specific animal model is required to simulate the manifestation of each clinical symptom. For example, *nof*-zebrafish mutants were used to simulate partial color blindness; Japanese medaka (*Oryzias latipes*) mutants chemically obtained through large-scale mutational screening were used to simulate hearing loss [14]. In their turn, N. Homma *et al*. used CRISPR/Cas9 to disrupt long-wavelength-sensitive opsins of medaka and produced color-blind fish [15]. In addition, the Japanese medaka fish serves as a model for studying the organogenesis process, hemoglobin-bilirubin metabolism, and many other processes.

To simulate Niemann–Pick disease type CI (NPC1), an inherited autosomal recessive lysosomal storage disorder, Tseng *et al.* used CRISPR/Cas9-mediated gene targeting to generate two npc1-null mutants. Since NPC1 is characterized by abnormal accumulation of unesterified cholesterol and glycolipids in late endosomes and lysosomes, and common signs include neonatal jaundice, hepatosplenomegaly, seizures, cerebellar ataxia and cognitive decline, one of the models showed early liver damage, and the second one manifested a neurological phenotype. CRISPR/Cas9-mediated gene targeting was used to induce double-strand DNA breaks: Wild-type zebrafish embryos (F0) were injected with npc1-specific guide RNA (sgRNA) and Cas9-encoding mRNA at the one-cell stage of development. The resulting F0 zebrafish were raised to adulthood and outcrossed to individual wild-type adults to obtain F1 embryos. The F1 embryos were screened for germline transmission of npc1 mutations by PCR and fragment analysis. In their turn, F0 adults carrying potential mutations in their germline were outcrossed to individual wild-type adults. F1 embryos obtained from this second outcross were raised to adulthood and screened for npc1 mutations. As a result, two mutant alleles, *npc1* y535 and *npc1* hg37 were identified. The observation revealed growth retardation and premature lethality in *npc1* mutants, while the histology confirmed high levels of unesterified cholesterol in the liver of fish larvae [16].

L.R. Strachan *et al*. were able to develop a fairly accurate zebrafish model of X-linked adrenoleukodystrophy (ALD). This disease belongs to the group of peroxisomal disorders with an X-linked recessive inheritance and is caused by mutations in the abcd1 gene, affecting peripheral and central nervous system myelin as well as adrenal glands. The *abcd1* gene encodes the transmembrane ALDP protein, which is required for the metabolism of very-long-chain fatty acids. It was found that the amino acid sequence of the zebrafish *abcd1* gene is 70% identical to the sequence of the human *abcd1* gene and is expressed in homologous regions in ontogeny. The TALEN technology was used to generate zebrafish abcd1 mutant allele lines and to induce mutations in exon 1, which produced premature stop codons. Observations showed that the abcd1 mutants had elevated levels of very-long-chain fatty acids in the body, development of hypomyelination, decreased survival, impaired motor function, and delayed development of the interrenal gland (the equivalent of the adrenal glands) [17].

5.2. Rodents

The biology specifics of mice have turned them into a mainstay of models used in studying a wide variety of pathological conditions and processes, even though the biology of small rodents differs significantly from the biology of larger animal species, let alone humans. The main problem is that the lifespan of mouse-like rodents is short; therefore, large animal models complement studies on the efficacy of gene therapy. Thus, large animal models of human genetic diseases complement murine studies, as they have a longer lifespan, they are more similar in size to a human neonate or a child, their background genetic heterogeneity is similar to that of humans, and they are genetically more closely related to humans than mice.

Nevertheless, mice and rats are inexpensive; they are quick to reach puberty; they have large litters. They are cheap and easy to maintain, and, due to their small size, they can be kept in large numbers in a relatively small area. Availability of murine models and an extensive knowledge base in genetics and immunology allow testing the effects of genetic manipulation and experimental therapeutics on physiological and pathological processes occurring in model organisms. In their turn, rats are much larger than mice and therefore, much better suited for manipulations and observation of the pathological process.

Leptin, a peptide hormone, belongs to the adipokines and regulates lipid metabolism, producing an anorexigenic effect; thus, a decrease in leptin levels leads to obesity. In 2017, Y. Chen *et al*. presented a method of obtaining a rat model of obesity, based on Sprague Dawley rats. The leptin receptor (Lepr) was knocked out through direct injection of *in vitro* transcribed mRNAs of TALEN pairs into rat zygotes to induce somatic mutations in 2 of 3 obtained pups, which were efficiently transmitted to the progenies [18].

T. Larcher *et al*. generated Dmd^{MDX} rats through TALENs targeting exon 23 of the *Dmd* gene to study Duchenne muscular dystrophy (DMD) [19]. These modified rats showed a significant decrease in muscle strength and a decrease in spontaneous

motor activity. Dmd^{MDX} rats may become a new fit-for-purpose small animal model of DMD.

Murine models are consistently used in studies of some clinical manifestations of cystic fibrosis, neurological and oncological diseases, severe combined immunodeficiency, hemophilia, DMD, Huntington's disease, β-hemoglobinopathies, metabolic diseases, and many other diseases waiting for gene therapy.

The experimental gene therapy of inner ear diseases (hearing loss, vestibular dysfunction) in mouse mutants demonstrated the possibility of partial restoration of lost functions by gene replacement, gene augmentation, and gene knockout and knockdown [20]. K.J. Carroll *et al.* have developed a transgenic murine model by injecting Cas9-encoding plasmid DNAs into mouse zygotes; the expression of Cas9 was regulated by the Myh6 promoter. In this transgenic model, high levels of Cas9 were expressed exclusively in heart cardiomyocytes. Then, the researchers delivered *sgRNAs* targeting *Myh6* by using an adeno-associated virus (AAV) vector, subsequently inducing cardiac-specific gene modification. The modification led to hypertrophic cardiomyopathy in transgenic mice [21]. By using TALENs and inducing a missense mutation at the *Scn8a* loci, J.M. Jones *et al*. have generated a murine model of early-onset epileptic encephalopathy. The data showed that this murine model would be useful for studying the development of pathogenesis and therapy of onset seizure disorders [22].

In 2014, a group of researchers led by W. Xue described a new technique of developing a liver cancer model in wild-type mice by using the CRISPR/Cas system *in vivo* and directly targeting tumor suppressor genes or inducing mutations of oncogenes in the liver. For this purpose, plasmid DNAs expressing Cas9 and *sgRNA* targeting the Pten and p53 tumor suppressor genes were hydrodynamically injected in combination and separately directly into the liver (**Fig. 5.3**) [23].

In the same year, R.J. Platt *et al.* demonstrated direct genome editing by AAV-mediated expression of Cre and sgRNA in the brain of Cre-dependent transgenic knock-in Cas9 mice. 3 weeks after the injection, deep sequencing of the *NeuN* locus showed that the Cas9 protein was functional and facilitated on-target indel

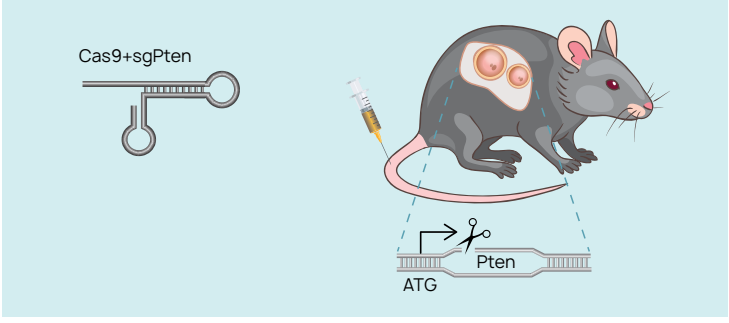


Fig. 5.3. Hydrodynamic injection of pX330 plasmids expressing Cas9 and sgRNA targeting the *Pten* gene sequence in wild-type mice to induce transient expression of the CRISPR/Cas components in hepatocytes

formation. Based on the results, the same group of researchers performed an AAV delivery of U6-sgRNA cassettes for the oncogene KRAS and the tumor suppressors p53 and LKB1 (AAV-KPL). Two months after the virus inoculation, the computed tomography found that all AAV-KPL-treated mice developed nodules in their lungs; the average total tumor burden over two months was approximately 10% of the total lung volume. The histological and genetic studies of these tumors showed that they were lung adenocarcinomas resulting from multiple alveolar adenomas, which developed within the first month after the inoculation of the vector construct [24].

Humanized models — transgenic animals containing functional genes, cells, tissues or other morphological structures of the human body — are especially important for pharmacological and toxicological studies. Murine models carrying a mutant or human β-globin gene made it possible to gain a better understanding of hemoglobin regulation and development mechanism of sickle cell disease or β-thalassemia [25].

Initially, studies focused on humanized monoclonal antibodies obtained *in vivo* and characterized by higher therapeutic efficacy compared to original antibodies of rodents [26, 27]. Homozygous humanized mice were developed in the late 20th century through the replacement of endogenous genes in the constant (C) region of the kappa light chain with human C kappa genes. Moreover, the resulting mice produced normal concentrations of serum antibodies, most of which carried chimeric kappa light chains and displayed normal immune responses [26]. In 2014, the large-scale *in situ* genetic humanization of mouse immunoglobulin genes was reported. The procedure was performed by using specially designed BAC-based recombinant targeting vectors combining human and mouse BACs, and by pronuclear injection into blastocyst or eight-cell morula stage mouse embryos. As a result, 6 million base pairs of mouse immune genes were replaced precisely in situ with their human orthologs; functionally and morphologically, the immune system of humanized mice was indistinguishable from that of wild-type mice [28]. The results of this study opened up the possibility for the humanization of other large complex loci, for example, major histocompatibility complex loci or T-cell receptor loci [29].

Such animals are especially important for modeling infectious diseases (**Fig. 5.4**) when the pathogen cannot be reproduced or the human-like immune response cannot be received in rodents.

One or a few genes inserted in a mouse is sufficient to replicate pathogens of such diseases as measles or Middle East respiratory syndrome coronavirus (MERS-CoV). Thus, by expressing the full-length human receptor dipeptidyl peptidase 4 (DPP4) protein or CRISPR/Cas-mediated amino acid alterations in the mouse DPP4 molecule, the mouse was made permissive to MERS-CoV replication [30, 31].

Another report informed about generating a fully immunocompetent murine model of Zika virus infection by using the mouse-adapted virus and replacing mouse stat2 with human stat2 (**Fig. 5.5**) [32].

Despite these advances in genome editing, some viral pathogens require a variety of genes in the host to facilitate the introduction and replication of virions and to induce the disease. Moreover, you should not forget about the limitations

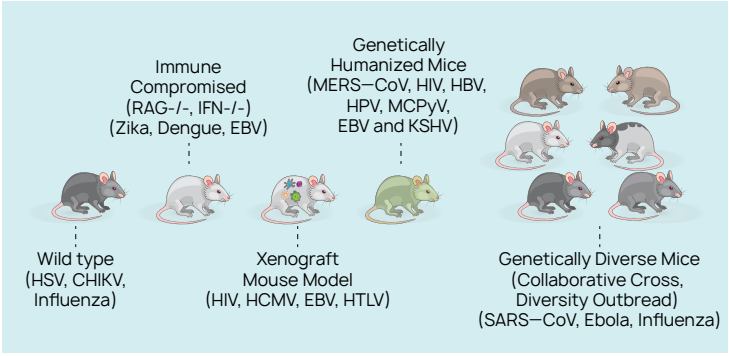


Fig. 5.4. Examples of murine models used in studies of human infectious diseases

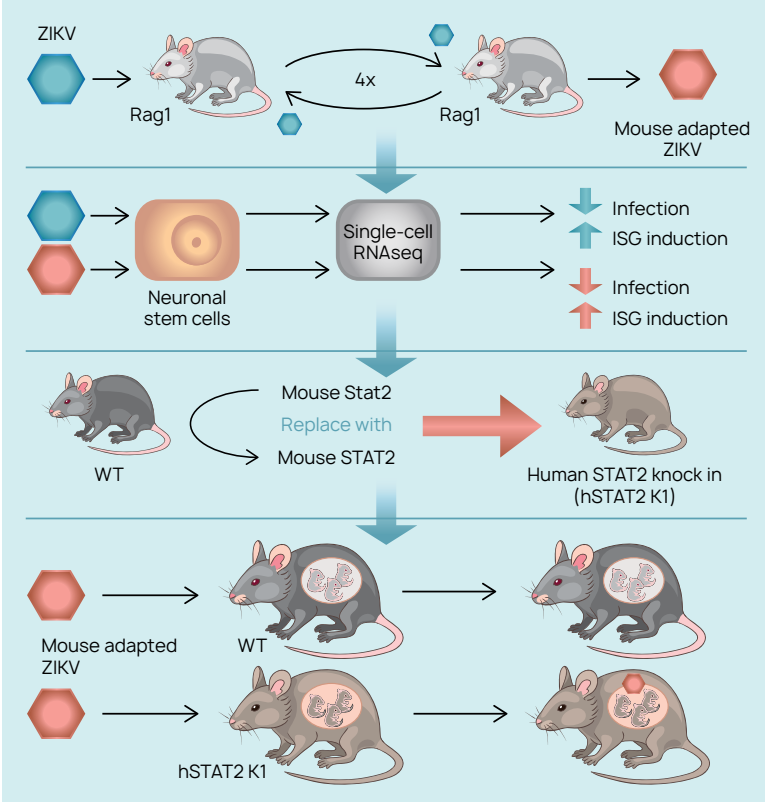


Fig. 5.5. Generation of an immunocompetent murine model of Zika virus infection

imposed by differences in the immune system of humans and rodents, when using standard or genetically humanized mouse strains. To overcome these complications, which impede the modeling of the human response to the pathogen, human tissue

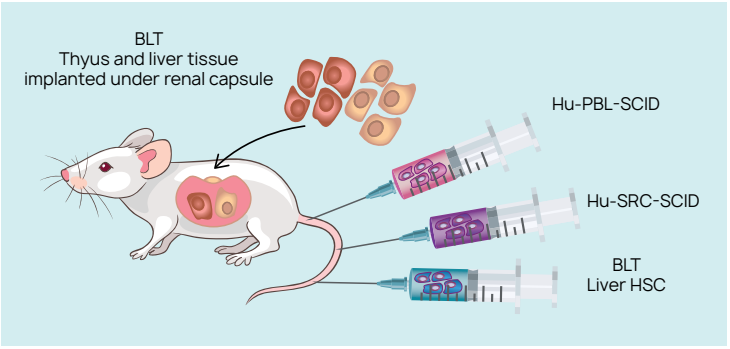


Fig. 5.6. Three ways to human immune system engraftment into an immunodeficient mice

mouse xenograft models are used. These models are based on mouse strains with a functionally absent adaptive immune system (for example, knockout mouse strains lacking *rag-1* or *IL2RG* genes). These mice are permissive to engraftment with human immune cells if some other mouse genes are removed (for example, the interleukin (IL)-2 receptor common gamma chain gene). Mice with xenografted human immune systems are used to study therapies for a wide range of diseases, including malaria, dengue, tuberculosis, and influenza. Simulating the processes occurring in the human body, these models have been especially useful in studying the pathogenesis of HIV and effectiveness of antiretroviral HIV therapy. Before CB17 immunodeficient mice were engrafted with human immune cells in 1988, the chimpanzee had been the only model for HIV research [33].

Currently, there are three main methods of introducing the human immune system into immunodeficient mice (**Fig. 5.6**).

The first model known as the Hu-PBL-SCID model was developed by intravenous injection of human peripheral blood leukocytes. This model is perfect for studying the function of human T cells *in vivo*, because it followed by the engraftment of human CD3+T cells during 7 days after the injection. It has a short experimental window, as mice develop lethal xenogeneic disease within 4–8 weeks.

The second model referred to as the humanized SCID repopulating cell (Hu-SRC)-SCID model is developed by an intrafemoral or intravenous injection of CD34+ hematopoietic stem cells obtained from bone marrow, cord blood, fetal liver, or peripheral blood. This model supports the engraftment of the entire human immune system by demonstrating the presence of B cells, T cells, myeloid cells, and antigen-presenting cells in peripheral hematopoietic tissues. However, granulocytes, erythrocytes, and platelets formed in the bone marrow are detected in whole blood only in small amounts. The BLT model represents the third version of introducing the human immune system into the mouse. The model was developed through transplantation of human fetal liver and thymus cells under the mouse kidney capsule in combination with intravenous injection of autologous fetal liver hematopoietic stem cells. Similar to the Hu-SRC-SCID model, the BLT model develops all human hematopoietic cell lines and a healthy mucosal immune system. In addition, T cells

are produced in the thymus autologous to the human one and are human leukocyte antigen restricted. Yet, this model also has a limited window for experiments, as mice develop the *graft-versus-host disease* (GVHD) syndrome [34].

Despite their usefulness, animal models with human xenografts have limitations, such as the efficiency of engraftment and tissue rejection as well as impossibility to fully predict human immune responses such as antigen-specific antibody responses. However, these models were a breakthrough in the study of plague pathogenesis and therapy in the 20th century.

Transgenic animals carrying highly differentiated human hepatocytes were required to simulate infection with the human hepatitis B virus, as normal mice were immune to HBV (cccDNA was not formed in mouse hepatocytes). Back in the mid-1980s, the first HBV-transgenic mice could only be used to study the role of HBsAg, HBeAg, and HBx proteins and their oncogenic potential. Although the obtained data were controversial, they set the stage for further HBV research.

In 1995, Chisari *et al*. demonstrated the ability of human HBV virions to replicate in mouse hepatocytes. At the same time, no damage to mouse hepatocytes was observed, indicating that HBV does not have cytopathic properties. These murine models were produced with a terminally redundant viral DNA construct that starts just upstream of the HBV enhancer, extends completely around the viral genome, and ends just downstream of the unique polyadenylation site in HBV [35]. Several years earlier, it was demonstrated that the adoptive transfer of cytotoxic T lymphocytes to transgenic mice induced acute hepatitis [36]. Thus, HBV transgenic mice have proven to be useful for testing antiviral agents that interfere with viral replication. To study the mechanisms of HBV pathogenesis and viral clearance, murine models have been developed through transduction of the virus genome into mice by using viral vectors or through hydrodynamic injection of the HBV genome.

Adenovirus or adeno-associated virus vectors containing HBV genomes transduce hepatocytes with high efficiency. However, the application of adenoviral vectors in mice triggers a vigorous immune response to the vector. This disadvantage was partially compensated by using adeno-associated viral vectors, though it resulted in mice resistance to viral antigens and persistent viremia in mice without HBV-specific humoral immune response [37].

When using the hydrodynamic injection technique, a large amount of native DNA solution is rapidly injected into mice through the tail vein, resulting in DNA uptake in hepatocytes and further transient gene expression. Although this technique has certain disadvantages, which can affect the host gene expression or signaling pathways, its advantage is that various genotypes and HBV variants or mutants can be injected into mice and analyzed *in vivo* within a relatively short time. In addition, the hydrodynamic injection opened the door for research in immune responses in the acute form of infectious diseases.

The discovery of the CRISPR/Cas 9 system gave hope for its further use in the selective removal of cccDNA. However, since the presence of cccDNA in mouse hepatocytes has not been detected, the targeting of CRISPR/Cas9 nucleases at cccDNA cannot be tested. Moreover, the main concern is the disruption of HBV sequences, which

can lead to genome instability and the appearance of tumors. In general, none of the above models stand up to criticism when judged from the perspective of gene therapy interventions. Therefore, humanized mice were developed with human hepatocytes stably integrated into the mouse liver parenchyma to achieve the long-term survival of highly differentiated human hepatocytes permissive to HBV infection *in vivo*.

Currently, three models of human liver chimeric mice are available to study hepatotropic virus infections. The Alb-uPA murine model is based on the property of the uPA (urokinase-type plasminogen activator) transgene to overexpress under the control of the albumin promoter. The hepatic uPA transgene overexpression results in elevated uPA plasma levels and leads to hypofibrinogenemia and subacute liver failure in neonatal transgenic mice. To prevent rejection of xenogeneic hepatocyte transplants, uPA mice were backcrossed to mice with genetic immunodeficiency ($Rag2^{-/-}$; Scid). The chimerism levels were assessed by measuring the human albumin serum levels in the blood of mice [38].

Another humanized model is based on mice deficient in fumarylacetoacetate hydrolase [39].

The third murine model is based on TK-NOG mice, which carry the herpes simplex virus thymidine kinase (TK) transgene under the control of the liver-specific albumin promoter and a triple immune defect, as they were backcrossed to NOG mice. The downside of this model is the infertility of males of this strain [40].

5.3. Rabbits

Unlike mice, the development of transgenic rabbits as a model was impeded for a long time by the low efficiency of embryonic stem cell usage and somatic cell nuclear transfer. Only in 2011, the first genetically modified rabbit with a disrupted endogenous IgM locus was produced by using the ZFN technology [41]. However, already in 2017, *Rag1, Rag2* [42], and *Fah* [43] knockout animals were obtained by using the TALEN technology. The CRISPR/Cas9 technology made it possible to develop rabbit models with a knockout of single or multiple genes at the same time; the probability of simultaneous multiple gene knockouts reached approximately 100% [44]. Recently, T. Sui *et al*. developed a DMD rabbit model by co-injecting Cas9 mRNA and two gRNAs into rabbit zygotes. The CRISPR system was intended to modify DMD exon 51, and DMD knockout rabbits exhibit the typical phenotypes of DMD. Moreover, clear pathology was also observed in the diaphragm and heart, similar to DMD patients. This new model may be more valuable for preclinical trials than the previous rodent models [45].

Thanks to rabbits, the results of other studies suggest that GJA8 gene encoding gap junction protein 50 was associated with autosomal-dominant congenital cataract [46, 47]. L. Yuan *et al*. co-injected Cas9-encoding mRNA and sgRNA into rabbit zygote to construct a GJA8 knockout rabbit model [48]. As a result, the gene mutation efficiency of the GJA8 site reached 98.7% and 100% in embryos and young rabbit tissues, respectively. They achieved efficient gene editing of the rabbit genome through the CRISPR/Cas9 system and provided a good disease model for cataract-related research.

D. Zhao *et al.* have published interesting results related to the use of gene therapy for the regeneration of bone defects. Synergistic effect of the recombinant plasmid pcDNA3.1-VEGF 165, which encodes vascular endothelial growth factor, and a gelatin sponge was demonstrated. The post-therapeutic histological examination showed the presence of a large amount of newly formed blood vessels two weeks after the beginning of treatment, a number of bone trabeculae with osteoblasts proliferation after four weeks, fresh periosteum and a reformed medullary cavity after 12 weeks, while in the control group these structures were formed in later phases [49].

Wilson disease is an autosomal recessive hereditary disorder of copper metabolism, which is caused by sequence variations in the *ATP7B* gene. ATP7B is an important protein contributing to the trans-membrane transport of copper. Recently, based on the CRISPR/Cas9 mediated single amino acid substitution, W. Jiang *et al*. have developed a rabbit Wilson disease model. At the onset of the disease, the copper content in the livers of modified rabbits increased nine-fold compared to wild-type rabbits; however, the survival rate of the models was approximately three months. Therefore, this model can be seen as a potential Wilson disease model for application in pathological analysis, clinical treatment, and evaluation of gene therapy efficacy [50].

Given that the gene sequences of rabbits have greater homology with human genes than rodents, and many diseases are comparable to those found in humans, the use of rabbits as an animal model puts them in a more advantageous position as compared to mouse-like rodents [51].

5.4. Pigs

Pigs are physiologically, anatomically, and genetically similar to humans. Therefore, they seem to be an ideal model for studying cardiovascular structure. In 2011, the conducted research combined ZFNs with the somatic cell nuclear transfer technology to develop pigs with a mutation in the γ-receptor activated by proliferator peroxisome (Ppar-γ). The PPAR-γ knock-out pig model served as a useful tool to study the role of PPAR-γ in cardiovascular diseases [52].

Marfan syndrome is an autosomal-dominant connective tissue disorder caused by mutations in the *FBN1* gene and displaying symptoms of cardiovascular and skeletal abnormalities. K. Umeyama *et al*. successfully obtained *FBN1*-mutant pigs by using ZFNs. The phenotypes of obtained animals, such as scoliosis, delayed mineralization of the epiphysis, funnel chest, and disrupted structure of elastic fibers of the aortic medial tissue, indicate the value of *FBN1*-mutant pigs as a model for better understanding the pathogenesis of Marfan syndrome and for developing further treatment [53].

For a better understanding of the mechanisms of Huntington's disease, in 2018, S. Yan *et al*. applied CRISPR/Cas9 technology to accurately insert human Huntington's mutation gene containing a 150CAG repeat into the endogenous pig *htt* gene locus. This resulted in the development of modified knock-in pigs expressing a human mutant huntingtin gene (*mHTT*) by somatic cell nuclear transfer (**Fig. 5.7**).

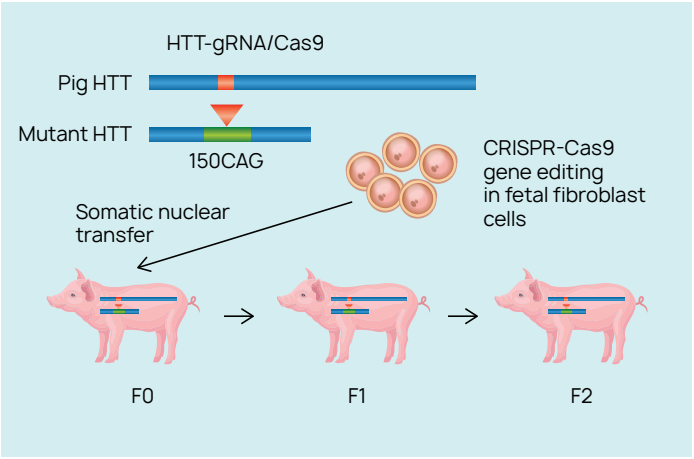


Fig. 5.7. Knock-in pigs developed by CRISPR/Cas edited fetal fibroblast nuclei transfer for studies in Huntington's disease

Moreover, the mutation was heritable. It was the world's first large animal model to simulate genetic mutations in patients with neurodegenerative diseases [54].

An important breakthrough in the study of neurodegenerative diseases is the development of pigs capable of generating the *htt* gene, which could contribute to the development of new drugs for the treatment of neurodegenerative diseases.

Genome editing technologies make it possible to develop models of metabolic diseases. The pathogenesis of one of them — familial hypercholesterolemia — stems from apolipoprotein E (*ApoE*) and low-density lipoprotein receptor (*LDLR*) gene defects, which can cause atherosclerosis. In 2012, Carlson *et al.* published the results of cloning TALEN-modified pigs. Cloning was carried out by chromatin transfer. The pregnancy of six sows resulted in 18 viable clones, eight of which contained monoallelic mutations, and ten — biallelic modifications of the *LDLR* gene [55]. A few years later, when the CRISPR/CAS technology gained popularity, L. Huang *et al*. reported the development of biallelic knockout pigs lacking *ApoE* and *LDLR* genes. Although ApoE–/–/LDLR–/– mice have been widely used in atherosclerosis research, studies have shown that lipoprotein profiles and metabolism in mice, in addition to their inability to reproduce other important signs of atherosclerosis, differ from those in humans and pigs. Biochemical studies of the blood serum of the resulting knockout piglets confirmed an increase in total cholesterol by almost 57%, and an increase in triglycerides by 120%, and these characteristics remained stable for a long time [56].

5.5. Cattle

Cattle are rarely used in gene therapy experiments because of their size and, consequently, the need for large quantities of recombinant protein preparations or for scaling up viral vectors. On the other hand, their size and weight make them suitable for

modeling pediatric diseases. Therefore, the main types of diseases that can be simulated in cattle are diseases of the musculoskeletal system and disorders of the urea cycle.

One of these diseases is citrullinemia, which is caused by mutations in the *ASS1* gene or in the *SLC25A13* gene. Citrullinemia is the main cause of hyperammonemia in children and leads to serious neurological disorders and death. In cattle, citrullinemia is caused by a nonsense mutation in the gene encoding arginosuccinate synthetase. To study this defect in the urea cycle, neonatal calves, being the most suitable models for growing children, received an adenoviral vector containing an intact copy of the arginosuccinate synthetase gene. A week after the administration of the adenovirus, the calves' serum showed elevated arginosuccinate synthetase levels, while all other liver parameters remained normal [57].

5.6. Small ruminants

The weight and size of sheep and goats tend to be close to those of human adolescents. However, due to the specifics of their physiology as well as the high cost of their maintenance, their use as models of human diseases is limited, although the comparability of some sheep diseases with human counterparts has been proven [58]. Back in the 2000s, there were attempts to use sheep as a corneal transplant model for humans. Since the corneal endothelium is the major target in human corneal graft rejection, the tested gene therapy options included transduction with herpes simplex virus type I (HSV-I) and an adenoviral vector encoding the *E. coli lacZ* reporter gene. The study showed that the HSV-I vector was not able to efficiently transduce the endothelium of the sheep cornea, while the adenoviral vector showed 70% transduction efficiency [59].

In addition, sheep are the only experimental model of Tay-Sachs disease (TSD), a lysosomal storage disease caused by a mutation in the *HEXA* gene. The *HEXA* gene encodes the α-subunit of the enzyme hexosaminidase A (HexA), the deficiency of which causes damage to the central nervous system. The intracranial gene therapy on TSD-sheep was tested by using AAVrh8 monocistronic vectors encoding the α-subunit of Hex or a mixture of two vectors encoding both the α and β subunits separately injected at high or low doses. The delay of symptom onset and/or reduction of acquired symptoms were noted in all adeno-associated virus-treated sheep. However, better HexA and vector genome distribution was achieved in the brain of animal models, when the two vectors were combined, although the distribution in the spinal cord remained low in all groups [60].

Another rare lysosomal storage disease is glycogen storage disease type V (also known as McArdle disease), which is characterized by muscle deformations, muscle pain and fatigue during exercise due to the buildup of abnormal amounts of glycogen in muscle tissue; the abnormal accumulation of glycogen is caused by the impaired catalytic function of muscle phosphorylase (myophosphorylase). By using the adenoviral vector AdV5, the adeno-associated vector AAV2, and plasmid DNAs, gene therapy was conducted to deliver DNA encoding myophosphorylase and the *LacZ* reporter gene (encodes the β-galactosidase enzyme) into the muscles

of affected sheep. Plasmids were delivered both by electroporation or sonoporation and without them. The application of viral vectors produced an increased number of transduced muscle fibers, which remained especially long after AdV5, while plasmids delivered by electroporation resulted in a higher level of transfection compared to other plasmids [61].

5.7. Horses

Osteoarthritis is the most common and economically significant disease in horses and humans. Using an established model of equine osteoarthritis that mimics clinical osteoarthritis, D.D. Frisbie *et al*. investigated the therapeutic effects resulting from intra-articular overexpression of the equine interleukin-1 receptor antagonist (IL-1 Ra) gene through adenoviral-mediated gene transfer. At the end of the study, clinical examination of the horses showed that the therapeutic expression of IL-1Ra significantly decreased signs of joint pain as measured by the degree of lameness [62].

L.M. Heinzerling *et al*. [63] studied gene therapy of another disease — melanoma. According to statistics, this disease affects up to 80% of grey horses over 12 years old. At the same time, the form and course of the disease have little in common with those in humans, but the histological and immunohistochemical features demonstrate many similarities with human neoplasms. This made it possible to use the grey horse as an adequate model in gene therapy of melanoma. The research results showed that injection of plasmid DNA encoding human interleukin-12 into melanoma metastases caused significant regression in all lesions. Moreover, one of the treated foci demonstrated a complete disappearance of metastases with no relapse after six months. During the treatment and in the post-treatment period, no side effects were observed in any of the animals [63].

5.8. Cats

Cats are believed less likely to suffer from hereditary diseases, but the specifics of cats' anatomy and physiology made it possible to choose them as adequate models for studying visual organs and the nervous system. Cats have large eyes, and their brain is anatomically similar to that of humans. This makes them most attractive for modeling diseases such as lysosomal storage diseases and spontaneous retinal diseases. The fact that some cats are a kind of ready-to-use models for some eye diseases has become known through genetic mapping.

Retinal degeneration was studied in cats of the Abyssinian breed — the rdAc model and the Rdy model [64]. Unfortunately, there is currently no information about the beginning of the genetic treatment of this disease.

On the other hand, there are constant reports of successful experimental therapy of various lysosomal storage diseases. Congenital α-mannosidosis is caused by a deficiency or defective function of the α-mannosidase enzyme, which leads to the accumulation of mannose-rich oligosaccharides in lysosomes. The main symptoms of this disease in humans are intellectual disability, ataxia, hepatosplenomegaly, gingival

hyperplasia, hearing loss, skeletal pathology, *etc*. In cats of the Persian breed, this disease can be inherited and is manifested in progressive cerebellar ataxia, polycystic kidney disease, hepatomegaly, skeletal anomalies, and other clinical signs similar to humans. The results of a series of studies in cats look very encouraging. Six injections of adeno-associated virus carrying the correct copy of the α-mannosidase gene were made into each rostral hemisphere of the brain and rostral brainstem; two injections were made into the cerebellum. The treatment resulted in reduced severity of cerebellar dysfunction signs. On the other hand, the control group of animals that did not receive treatment showed aggravation of cerebellar dysfunction. The analysis of the lysosomal stores also showed positive dynamics, although they were increasing in numbers with the increased distance from the injection site. None of the brain regions of the treated cats showed the amount of lysosomal stores comparable to the large amount in control animals [65].

Experimental gene therapy of GM1-gangliosidosis was performed in GM1 feline models by delivering an AAV vector expressing β-Gal (β-galactosidase) and bilateral injection into the thalamus and deep cerebellar nuclei. The long-term follow-up of GM1 cats showed a statistically significant five-fold increase in survival compared to GM1 models not receiving therapeutic agents [66]. The feline model of GM2 gangliosidosis caused by one of the four mutations in the HEXB gene resulting in both HexA and HexB enzyme deficiencies is a true model of human GM2 gangliosidosis (Sandhoff disease). Therapy of similar models was also performed by using bilateral intrathalamic injections of AAV vectors encoding α and β-Hex subunits. At first, it led to the doubling of the animals' lifespan and after the method was modified, the lifespan of treated animals increased four times compared with the control group [65].

Mucopolysaccharidosis Type I (MPSI) in humans is characterized by a deficiency of the lysosomal enzyme α-L-iduronidase (IDUA; an enzyme that hydrolyzes the terminal α-L-iduronic acid residues of two glycosaminoglycans — dermatan sulfate and heparan sulfate), which entails the accumulation of partially decomposed dermatan sulfate and heparan sulfate in lysosomes. It results in clinical symptoms most frequently combining retarded intellectual and physical development, corneal opacity, organomegaly, high levels of glycosaminoglycans in the urine, *etc*. In terms of symptoms, the feline model of MPSI is closest to severe MPSI in humans. The successful cloning of the feline *IDUA (fIDUA)* cDNA sequence by L.M. Hinderer *et al*. made it possible to study gene therapy not only in canine models but also in feline models. Early studies in cats were carried out by using the canine *IDUA* genetic product, which caused the development of a cytotoxic T-lymphocytic immune response. The consolidated results of therapeutic intrathecal delivery of *AAV9*-fIDUA showed global CNS transduction, normalization of secondary lysosomal enzymes, and reduction in damages associated with the accumulation of glycosaminoglycans, cholesterol, and GM3-ganglioside [67], despite an immune response. Further studies of the intravenous administration of AAV9-fIDUA showed complete correction of MPSI-associated cardiovascular lesions in feline models [68].

Speaking about infectious diseases, special attention should be given to the feline model of HIV. All members of the feline family are susceptible to their immunodeficiency disease (FIV) caused by a lentivirus from the *Retroviridae* family and leading to progressive depletion of the immune system and, ultimately, to acquired immunodeficiency syndrome. Structurally and sequentially, FIV is very close to HIV, which also belongs to the *Lentivirus* genus. Morphologically it is a spherical virion of 120 nm in diameter (compared to 100 nm in FIV) and containing a diploid set of genes — a pair of copies of a single-stranded positive-sense RNA packed together with nucleocapsid (p7) and additional proteins (reverse transcriptase, integrase, protease). By analogy with FIV, the dense core of the HIV virion formed by capsid protein (p24) contains a ribonucleoprotein complex surrounded by a spherical shell of a matrix protein (p17) [69]. Like FIV, HIV requires initial interaction with the primary binding receptor. However, unlike FIV (whose attachment receptor is represented by the surface molecule CD134, which allows infection of B cells and $CD8^+$-T cells in addition to $CD4^+$ lymphocytes, macrophages, or monocytes), HIV uses $CD4^+$ as the main binding receptor and CCR5 as the main entry receptor. Moreover, the *orfA* gene (encoding the OrfA protein), which was erroneously seen as FIV transcription transactivator, has a similar localization with the *Tat* gene encoding the HIV transactivator [70]. It was shown that FIV virions, like HIV, in the acute phase invade the central nervous system through infected lymphocytes and monocytes, or the free virus penetrates the blood-brain barrier. Once present in the CNS, both FIV and HIV infection spreads to microglia and astrocytes, which then serve as a reservoir for latent viral persistence [71].

Such characteristics caught the attention of researchers to this *natural* model for human immunodeficiency research; therefore, FIV-infected cats are used to study the pathogenesis of HIV, for example, neurological or immune dysfunction as well as HIV-associated diseases [71]. The results of research on vaccines against FIV and the prospects for obtaining data that would underlie the strategy for vaccination against HIV are controversial. It has been found that the developed vaccine does not provide immunity against some FIV strains. At the same time, other developed vaccines against FIV did not generate any protective immunity or led to increased susceptibility to the pathogen through antibody-dependent enhancement or general immune activation [72, 73]. Despite this, the FIV model has significant potential as a reliable tool for the assessment of the efficacy of new HIV therapies.

5.9. Dogs

Many immune system components in dogs are similar to those in humans. Moreover, dogs' size and lifespan are comparable to those of children. Therefore, it is not surprising that more than 50% of genetic diseases in dogs are caused by mutations in the same genes as in humans.

To assess the efficacy and safety of experimental gene therapy for MPS in canine models, scientists studied the effect of adeno-associated viral vectors, through which IDUA was administered to the brain. The researches were conducted on the known

models of MPS Type III B (Sanfilippo syndrome B) and on the above mentioned MPSI. During the research, it was found that immunosuppression was a major determinant of treatment efficacy; therefore, immunosuppression should be maintained in dogs. The absence of immunosuppressant in some MPSIIIB models resulted in low vector copy numbers, absence of detectable activity of N-acetyl-alpha-glucosaminidase (NAGLU) (the enzyme that degrades heparan sulfate by hydrolysis of terminal N-acetyl-D-glucosamine residues in N-acetyl-alpha-D-glucosaminides), almost unchanged pathology severity, and manifest inflammatory response [74].

The DMD originally identified in a young golden retriever is known as golden retriever muscular dystrophy (GRMD). Therefore, despite the presence of murine models, the GRMD model is a clinically more suitable model for studying this disease in humans. Current gene therapy studies in animal DMD models include plasmid DNA and viral vector therapeutic techniques. Direct intramuscular injections of a plasmid encoding human dystrophin generated several dystrophin-positive cells, while electrotransfer of plasmids encoding both the full-length dystrophin and microdystrophin caused limited expression and increased cellular infiltrates. The application of adenoviral (AdV) and AAV vectors brought positive results. However, there was a side effect: a primary cellular response against capsid and transgenic proteins; the problem was solved by inducing immune suppression in animals.

Promising data were obtained thanks to multi-exon skipping by using antisense nucleotides to restore the reading frame and obtain shorter but functional dystrophin proteins. Studies focusing on vector delivery of these RNAs via AAV or AAV-U7 to muscles have provided safety and efficacy data for preclinical trials through a dose-dependent response that increased dystrophin expression and decreased pathology in treated skeletal muscle [75].

The most common form of hemophilia is hemophilia A (HA) caused by a mutation of the coagulation factor VIII gene and resulting in factor VIII (FVIII) deficiency. Since the treatment with recombinant or plasma-derived FVIII proteins becomes ineffective over time due to the development of FVIII-specific antibodies and high costs of the treatment, gene therapy methods become attractive ways to maintain health in patients with hemophilia A.

The selection of dogs as a basis for modeling hemophilia was especially successful not only because they are physiologically close to children, but also because these models are natural (dogs were susceptible to three types of disease: A, B, and C).

Although the first steps in gene therapy of canine HA models were quite successful, the introduction of human cDNA encoding FVIII via AdV vectors was ineffective due to the short duration of expression [76] caused, according to some authors [77, 78], by two-phase toxicity of early generation vectors. Therefore, all subsequent researches were aimed at the reduction of toxicity in vector viruses. W.M. McCormack *et al*. reported that after the adenoviral HDV-PEPCK/BDD-cFVIII/WPRE vector was administered as a vehicle for canine FVIII (cFVIII) cDNA, they observed a transient dose-dependent elevation in liver enzymes and thrombocytopenia that disappeared within two weeks. At the same time, the parameters used to evaluate the effectiveness of therapy (whole blood coagulation time, concentration and activity

of cFVIII in plasma, activated partial thromboplastin time) demonstrated significant improvement in all experimental animals. In two animals receiving an increased dose, the whole blood coagulation time reached practically normal values, despite the decreasing level of cFVIII activity for two years. In addition, persistent vector-specific DNA and RNA were found in liver tissue, though antiviral antibodies were not found. The authors also note that after vein punctures, the dogs did not have such a clinical sign of hemophilia as prolonged bleeding [79]. Thus, it was shown that a severe form of hemophilia A could be converted into a mild form.

Gene therapy for the deficiency of another coagulation factor, proconvertin (FVII), has been successfully demonstrated in FVII-G96E canine models. The treatment was performed with escalating doses (from 2E11 to 4.95E13 vector genomes per kg body weight) of the AAV serotype 8 containing DNA complementary to cFVII for zymogen. The research results showed the prolonged efficiency and safety of the construction as well as the adequacy of FVII-G 96E dogs as models [80].

Human X-linked severe combined immunodeficiency (X-SCID) is associated with interleukin-2 receptor γ-chain mutations, which are fatal for children in the first years of their life due to the complete absence of cellular and humoral immunity. Two distinct spontaneous mutations have been identified in dogs, resulting in genuine clinical, pathological, and immunological X-SCID models. Hematopoietic stem cell transplantation without prior transplant conditioning has been recognized as the standard treatment for patients with X-SCID. A similar therapy was used in canine X-SCID models by using normal dogs as bone marrow donors. In contrast to X-SCID-affected individuals who had donor T-cell engraftment and T-cell function recovery, but poor donor B-cell engraftment and poor recovery of the humoral immune response, canine X-SCID models demonstrated 100% circulation of donor T cells and 20–50% of donor B cells as well as a complete restoration of the immune function [81]. Despite the decrease in T-cell diversity over time after the transplantation, X-SCID dogs could reach a survival limit of more than ten years [82].

An attempt to use genetically improved autologous hematopoietic stem cells as an alternative to classical hematopoietic stem cell transplantation via γ-retroviral gene therapy initially showed good results, but a gradual decrease in T-cell levels resulted in the death of all dogs within eleven months after the transplantation [83].

In vivo gene therapy by intravenous injection of RD114-pseudotyped retroviral vector demonstrated viral expression in peripheral blood lymphocytes three weeks after the injection, which increased and accounted for 85% of T cells with corrected genes eight weeks after the treatment. The long-term follow-up observation of dogs revealed a stable correction of T lymphocytes and constant circulation of up to 26% of genetically corrected B lymphocytes as well as the presence of the vector in myeloid lines and normalization of the immune system function up to 18 months after the end of therapy [84]. P.J. Felsburg *et al.* reported that the gene-correction of T cells in lentiviral treated dogs was sustained for up to 4.5 years [85]. The results of X-SCID gene therapy *in vivo* by using a foamy virus vector were published in 2014. It was reported that all dogs from the treatment group showed the presence of genetically modified lymphocytes two weeks after the injection. The number of

genetically modified lymphocytes continued to increase for 12 weeks. However, the survival range of XSCID models varied from three to 10.5 months [86].

Leber congenital amaurosis (LCA) accounts for about 15% of diseases associated with inherited congenital blindness. Leber amaurosis can be caused by the disappearance of the lecithin/retinol acyltransferase enzyme due to mutation in the RPE 65 gene. The enzyme is required for the regeneration of photopigment in retina. The natural animal RPE65$^{-/-}$ model has early and serious visual impairments similar to those of a human suffering from Leber congenital amaurosis. Therapy with a recombinant AAV carrying wild-type RPE65 yielded good results: improved vision allowed the dogs to pass through the labyrinth [87]. This made it possible to conduct successful researches in humans.

The research by S.R. Bianco *et al*. was based on the property of the Fas ligand to induce apoptosis of melanoma cells. Liposomes conjugated to Fas-ligand DNA were injected intralesionally and into the tissues surrounding the neoplasm. Further observations demonstrated a reduction in oral melanoma by 12–58% in three out of five dogs [88]. This model can be considered adequate since it is natural.

5.10. Non-human primates

Non-human primates (NHPs) have an innate superiority compared to other animal models due to their similarities to humans in genetics, physiology, developmental biology, social behavior, and cognition. NHPs can be an ideal model, especially in modelling nervous system diseases [89].

Although gene manipulation in monkeys is far more difficult than in other animals (it took more than 25 years of research to establish transgenic NHPs, after the first transgenic mouse was developed), the discovery of genome editing has contributed significantly to the progress of work on generating NHP models.

Rett syndrome (RTT) is an X-linked neurodevelopmental disorder in the autism spectrum. It is known that loss-of-function mutations of methyl-CpG-binding protein 2 (MECP2) lead to Rett syndrome. Y. Chen *et al.* designed three pairs of TALENs to target multiple sites on exon 3 of *MECP2*. All three TALEN pairs of targeted plasmids, either individually or in combination, were injected into single-cell monkey zygotes. The results showed that the *MECP2*-mutant males died during gestation, while the mutant females developed physiological and behavioral disorders. Notably, these disorders were similar to those of humans with RTT disease. This animal model has provided more opportunities for studying disease mechanisms and seeking treatment options [90]. In another study, the CRISPR/Cas9 system was used to target exons 4 and 46 of DMD to generate DMD monkey models [91].

Primary immunodeficiencies are represented by a diverse group of rare and chronic diseases. Part of the body's immune system is missing or functions improperly, threatening the lives of patients. Severe combined immunodeficiency (SCID) is the most severe form of primary immunodeficiencies [92].

In 2016, Japanese scientists optimized ZFNs and TALENs to generate indels at interleukin-2 receptor subunit γ-locus (IL2RG) in pronuclear stage monkey

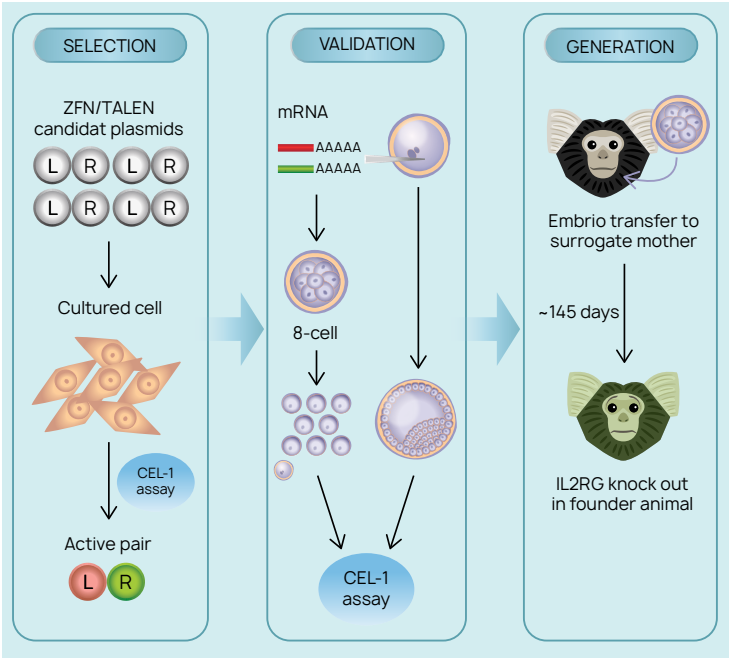


Fig. 5.8. Schematic representation of generating NHPs with the SCID phenotype by using ZFN and TALEN technologies

embryos. Detectable DNA mismatches at the target locus led to inactivation of IL2RG concomitant with immunodeficiency (**Fig. 5.8**). They demonstrated a highly effective generation of founder NHPs with SCID phenotypes [93].

Chimpanzees have always been a reliable model in HBV and HCV researches. However, in recent years, due to the tightening of ethical norms and rules, these studies are extremely limited. Therefore, scientists turned to NHPs and other animals to use them as models of these infectious diseases. Until recently, all attempts to transmit HBV to nonhuman primates were generally unsuccessful. The only exception was macaques who maintained HBV replication after intrahepatic HBV DNA inoculation. But even in this case, the maximum result is self-limited hepatitis [94]. Thus, further studies are needed to assess the suitability of this model for studying the antiviral and immunological aspects of HBV infection.

Speaking about the problem of HBV infection modeling, we cannot but mention animals that are not primates but are experimentally susceptible to HBV — tupaias (*Scandentia*). The low infection efficiency of standard laboratory animals, combined with ethical and experimental limitations, made the HBV tupaia model the most suitable for *in vivo* studies. It has been found that neonatal animals having HBV-induced chronic infection develop moderate viremia levels and liver immunopathology, including fibrosis and hepatocellular carcinoma [95].

It may seem that NHPs with their highly developed cerebral cortex, cognitive capabilities, complex motor functions, and human-like neuroanatomy would be perfect

models for age-related neurodegenerative disorders characterized by progressive neuronal cell death, such as Alzheimer's disease (AD), Parkinson's disease (PD), or Huntington's disease (HD). However, recent advances in the study of these disorders only slightly improved the understanding of the pathogenesis and treatment of some of the symptoms associated with these diseases.

Aging or targeted cholinergic lesions damage in the brain was common methods of simulating AD in primates. Age-related disorders observed in elderly primates had some signs of AD, but they did not include all aspects of the disease [96–99]. The cognitive deficit associated with the loss of cholinergic neurons were modeled by stereotaxic injections of the ibotenic acid into the basal forebrain [100], while the damage to the hippocampus and caudate nucleus resulted in suppression of the ability to learn complex tasks [101]. Recent studies have confirmed the involvement of nerve growth factor (NGF) in the maintenance of cholinergic neurons; however, NGF is unable to cross the blood-brain barrier; therefore, the treatment on models required targeted intracerebral NGF delivery, for example, by genetically modified autologous fibroblasts [102]. The efficiency of these methods paved the way for Phase I clinical trials in 2005, the positive results of which were published in 2015 [103]. AAV and lentiviral vectors have served as an alternative option for personalized cellular delivery of NGF. M.H. Tuszynski *et al*. demonstrated the ability of AAV-NGF to induce long-term biologically active NGF expression [103].

The first attempts to use primates and simulate movement disorders typical of HD were also limited to neurotoxin-mediated induction of lesions in striatum. Data describing the first NHP model of HD were published only in 2008. Gene transfer was performed by injection of rhesus monkey mature oocytes into the perivitelline space with high titer lentiviral vectors expressing exon 1 of the human HTT gene with 84 CAG repeats. All primates were delivered at full term and carried transgenic mutant HTT genes, although the length of the repeats ranged from 27 to 88, which resulted in the death of two animals during the first day, and one animal died at the age of one month [104]. As for the application of new generation gene therapy methods, they are currently being studied on cell cultures as well as in mouse and NHP models.

PD is a neurodegenerative disorder that is mainly caused by the degeneration of nigrostriatal dopaminergic neurons, and is characterized by such symptoms as tremor, rigidity, postural instability, sluggishness and abnormal gait [105]. As in previous cases, neurotoxins were very popular in simulating motor dysfunction, though their action could not reproduce the full picture due to their toxic effect on neurons, poorly mimicking the progress of the disease. Among the genes associated with PD, there are α-synuclein (*SCNA*) and leucine-rich repeat kinase 2 (*LRRK2*).

One of the pressing problems hindering the progress in the development of effective therapeutic options for the treatment of PD is the absence of a versatile animal model. The existing models are available in two options: acute (neurotoxin-induced) and chronic (genetic). Neurotoxic models are based on the administration of 6-hydroxydopamine [106] or 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine [107, 108] and the destructive effect of the resulting free radicals on dopaminergic neurons.

Like the neurotoxic models of the previous neurodegenerative diseases, acute PD models mimic only motor symptoms of the disease without progressive death of dopaminergic cells, thus significantly impeding the development of therapeutic strategies. The development of transgenic mice carrying missense mutations in the genes responsible for SNCA (α-synuclein) and LRRK2 (leucine-rich repeat kinase 2, also known as dardarin) made it possible to reproduce limited parkinsonism with moderate nigrostriatal degeneration of dopaminergic neurons [109–111]. To develop a chronic PD model, intracerebral injection of AAV vectors or lentiviral vectors containing mutated *SNCA* have been tested [112–114]. NHP models represented by cynomolgus monkeys (*Macaca fascicularis*) or rhesus monkeys (*Macaca mulatta*) demonstrated movement disorders and neuropathological features typical of PD.

The encouraging research data on the application of a helper-dependent canine adenovirus (CAV-2) as a delivery vector [115] makes it ideal for modeling PD in NHPs. In rodents and NHPs, these vectors preferentially transduce neurons; they have no long-term impact on newborn neuron homeostasis and have a cloned capacity of up to 30,000 bp. Recently, CAV-2 vectors have been generated; the vectors contain a $LRRK2^{G2019S}$ expression cassette, which was injected into a putamen of a rather interesting NHP — a gray mouse-like lemur (*Microcebus murinus*), which has positive traits of mouse-like rodents (small size, large litter sizes in captivity) and features of the nervous system of NHPs [116].

As for the gene therapy for PD, the treatment of patients is still ineffective due to the imperfection of animal models.

In 2018, a multidisciplinary team of veterinarians and researchers reported that a small population of Japanese macaques (*Macaca fuscata*) carried a mutation in the *CLN 7* gene, which causes one of the forms of Batten disease, a rare fatal neurodegenerative recessively inherited disorder that belongs to the group of neuronal ceroid lipofuscinoses. This discovery will make it possible to develop and test a strategy for Batten disease gene therapy since this macaque population is the only known model of this disease among NHPs in the world [117].

NHP models serve as an intermediate link for understanding the mechanisms of a disease and as a bridge connecting experimental treatment of simulated pathological conditions and therapy of truly suffering patients. Improvement of technologies for targeted genome editing and methods of embryonic engineering will make it possible to transfer NHPs to the category of routine laboratory animal models for studying gene therapy and drug screening.

The opportunity to introduce the reader to the diversity of existing animal models is limited by the size of our monograph. Researchers of the entire world community are making attempts to simulate not only individual symptoms (though it is justified at initial stages), but also pathogenesis leading to the manifestation of a complete symptom complex of human disease. Close cooperation of research institutes and laboratories with the global veterinary community is extremely important and essential because some diseases typical of animals can become adequate models for human diseases, and complete decoding of the genome, including wild animals, will reveal new adequate platforms for modeling.

Undoubtedly, modeling on one species of animals cannot fully demonstrate the entire pathway of a damaging agent, from its invasion to the development of a severe complex of multiple organ dysfunction, whether it is a pathogen or an inherited gene mutation. Therefore, it is quite natural that during the next years, neither zebrafish nor even flatworms (which, unfortunately, we left without due attention) will lose their relevance as models for the research of the pathogenesis of various diseases and for the development of strategies evaluating the efficiency of gene therapy through systems of targeted genome editing.

References

1. Blagbrough I.S., Zara C. Animal models for target diseases in gene therapy — using DNA and siRNA delivery strategies. *Pharm. Res.* 2009; 26(1): 1.
2. Simmons D. The use of animal models in studying genetic disease: transgenesis and induced mutation. *Nature Education.* 2008; 1(1): 70.
3. Rosen B.H., Chanson M., Gawenis L.R., et al. Animal and model systems for studying cystic fibrosis. *J. Cyst. Fibros.* 2018; 17(2S): S28-S34.
4. Hardouin S.N., Nagy A. Mouse models for human disease. *Clin. Genet.* 2000; 57(4): 237-44.
5. Bouabe H., Okkenhaug K. Gene targeting in mice: a review. *Methods Mol. Biol.* 2013; 1064: 315-36.
6. Jimenez N., Coll J., Salguero F.J., et al. Co-injection of interleukin 8 with the glycoprotein gene from viral haemorrhagic septicemia virus (VHSV) modulates the cytokine response in rainbow trout (Oncorhynchus mykiss). *Vaccine*. 2006; 24(27-28): 5615-26.
7. Rossi A., Kontarakis Z., Gerri C., et al. Genetic compensation induced by deleterious mutations but not gene knockdowns. *Nature*. 2015; 524(7564): 230-3.
8. Novodvorsky P., Watson O., Gray C., et al. klf2ash317 mutant zebrafish do not recapitulate morpholino-induced vascular and haematopoietic phenotypes. *PLoS One.* 2015; 10(10): e0141611.
9. Ota S., Hisano Y., Ikawa Y., et al. Multiple genome modifications by the CRISPR/Cas9 system in zebrafish. *Genes Cells.* 2014; 19(7): 555-64.
10. Kotani H., Taimatsu K., Ohga R., et al. Efficient multiple genome modifications induced by the crRNAs, tracrRNA and Cas9 protein complex in zebrafish. *PLoS One.* 2015; 10(5): e0128319.
11. Cao J., Navis A., Cox B.D., et al. Single epicardial cell transcriptome sequencing identifies Caveolin 1 as an essential factor in zebrafish heart regeneration. *Development*. 2016; 143(2): 232-43.
12. Zou J., Tran D., Baalbaki M., et al. An internal promoter underlies the difference in disease severity between N-and C-terminal truncation mutations of Titin in zebrafish. *Elife*. 2015; 4: e09406.
13. Kennedy B.N., Alvarez Y., Brockerhoff S.E., et al. Identification of a zebrafish cone photoreceptor–specific promoter and genetic rescue of achromatopsia in the nof mutant. *Invest. Ophthalmol. Vis. Sci.* 2007; 48(2): 522-9.
14. Furutani-Seiki M., Sasado T., Morinaga C., et al. A systematic genome-wide screen for mutations affecting organogenesis in Medaka, Oryzias latipes. *Mech. Dev.* 2004; 121(7-8): 647-58.
15. Homma N., Harada Y., Uchikawa T., et al. Protanopia (red color-blindness) in medaka: a simple system for producing color-blind fish and testing their spectral sensitivity. *BMC Genet.* 2017; 18(1): 10.
16. Tseng W.C., Loeb H.E., Pei W., et al. Modeling Niemann-Pick disease type C1 in zebrafish: a robust platform for in vivo screening of candidate therapeutic compounds. *Dis. Model. Mech.* 2018; 11(9): dmm034165.

17. Strachan L.R., Stevenson T.J., Freshner B., et al. A zebrafish model of X-linked adrenoleukodystrophy recapitulates key disease features and demonstrates a developmental requirement for abcd1 in oligodendrocyte patterning and myelination. *Hum. Mol. Genet.* 2017; 26(18): 3600-14.
18. Chen Y., Lu W., Gao N., et al. Generation of obese rat model by transcription activator-like effector nucleases targeting the leptin receptor gene. *Sci. China Life Sci.* 2017; 60(2): 152-7.
19. Larcher T., Lafoux A., Tesson L., et al. Characterization of dystrophin deficient rats: a new model for Duchenne muscular dystrophy. *PLoS One.* 2014; 9(10): e110371.
20. Wang L., Kempton J.B., Brigande J.V. Gene therapy in mouse models of deafness and balance dysfunction. *Front. Mol. Neurosci.* 2018; 11: 300.
21. Carroll K.J., Makarewich C.A., McAnally J., et al. A mouse model for adult cardiac-specific gene deletion with CRISPR/Cas9. *Proc. Natl. Acad. Sci. USA.* 2016; 113(2): 338-43.
22. Jones J.M., Meisler M.H. Modeling human epilepsy by TALEN targeting of mouse sodium channel Scn8a. *Genesis.* 2014; 52(2): 141-8.
23. Xue W., Chen S., Yin H., et al. CRISPR-mediated direct mutation of cancer genes in the mouse liver. *Nature*. 2014; 514(7522): 380-4.
24. Platt R.J., Chen S., Zhou Y., et al. CRISPR-Cas9 knockin mice for genome editing and cancer modeling. *Cell*. 2014; 159(2): 440-55.
25. McColl B., Vadolas J. Animal models of β-hemoglobinopathies: utility and limitations. *J. Blood Med.* 2016; 7: 263-74.
26. Zou Y.R., Gu H., Rajewsky K. Generation of a mouse strain that produces immunoglobulin kappa chains with human constant regions. *Science.* 1993; 262(5137): 1271-4.
27. Zou Y.R., Müller W., Gu H., et al. Cre-loxP-mediated gene replacement: a mouse strain producing humanized antibodies. *Curr. Biol*. 1994; 4(12): 1099-103.
28. Murphy A.J., Macdonald L.E., Stevens S., et al. Mice with megabase humanization of their immunoglobulin genes generate antibodies as efficiently as normal mice. *Proc. Natl. Acad. Sci. USA*. 2014; 111(14): 5153-8.
29. Macdonald L.E., Karow M., Stevens S., et al. Precise and in situ genetic humanization of 6 Mb of mouse immunoglobulin genes. *Proc. Natl. Acad. Sci. USA*. 2014; 111(14): 5147-52.
30. Li K., Wohlford-Lenane C.L., Channappanavar R., et al. Mouse-adapted MERS coronavirus causes lethal lung disease in human DPP4 knockin mice. *Proc. Natl. Acad. Sci. USA*. 2017; 114(15): E3119-28.
31. Cockrell A.S., Yount B.L., Scobey T., et al. A mouse model for MERS coronavirus-induced acute respiratory distress syndrome. *Nat. Microbiol.* 2016; 2(2): 16226.
32. Gorman M.J., Caine E.A., Zaitsev K., et al. An immunocompetent mouse model of Zika virus infection. *Cell Host Microbe.* 2018; 23(5): 672-85.e6.
33. Sarkar S., Heise M.T. Mouse models as resources for studying infectious diseases. *Clin. Ther.* 2019; 41(10): 1912-22.
34. Walsh N.C., Kenney L.L., Jangalwe S., et al. Humanized mouse models of clinical disease. *Annu. Rev. Pathol.* 2017; 12: 187-215.
35. Guidotti L.G., Matzke B., Schaller H., et al. High-level hepatitis B virus replication in transgenic mice. *J. Virol.* 1995; 69(10): 6158-69.
36. Ando K., Moriyama T., Guidotti L.G., et al. Mechanisms of class I restricted immunopathology. A transgenic mouse model of fulminant hepatitis. *J. Exp. Med.* 1993; 178(5): 1541-54.
37. Yang D., Liu L., Zhu D., et al. A mouse model for HBV immunotolerance and immunotherapy. *Cell Mol. Immunol.* 2014; 11(1): 71-8.

38. Dandri M., Petersen J. Chimeric mouse model of hepatitis B virus infection. *J. Hepatol.* 2012; 56(2): 493-5.
39. Azuma H., Paulk N., Ranade A., et al. Robust expansion of human hepatocytes in Fah$^{-/-}$/Rag2$^{-/-}$/$^{-/-}$Il2rg$^{-/-}$ mice. *Nat. Biotechnol.* 2007; 25(8): 903-10.
40. Sun S., Li J. Humanized chimeric mouse models of hepatitis B virus infection. *Int. J. Infect. Dis.* 2017; 59: 131-6.
41. Flisikowska T., Thorey I.S., Offner S., et al. Efficient immunoglobulin gene disruption and targeted replacement in rabbit using zinc finger nucleases. *PLoS One*. 2011; 6(6): e21045.
42. Song J., Zhong J., Guo X., et al. Generation of RAG 1-and 2-deficient rabbits by embryo microinjection of TALENs. *Cell Res.* 2013; 23(8): 1059-62.
43. Li L., Zhang Q., Yang H., et al. Fumarylacetoacetate hydrolase knock-out rabbit model for hereditary tyrosinemia type 1. *J. Biol. Chem.* 2017; 292(11): 4755-63.
44. Yan Q., Zhang Q., Yang H., et al. Generation of multi-gene knockout rabbits using the Cas9/gRNA system. *Cell Regen.* 2014; 3(1): 12.
45. Sui T., Lau Y.S., Liu D., et al. A novel rabbit model of Duchenne muscular dystrophy generated by CRISPR/Cas9. *Dis. Model Mech.* 2018; 11(6): dmm032201.
46. Zhu Y., Yu H., Wang W., et al. A novel GJA8 mutation (p. V44A) causing autosomal dominant congenital cataract. *PLoS One.* 2014; 9(12): e115406.
47. Jiang J.X. Gap junctions or hemichannel-dependent and independent roles of connexins in cataractogenesis and lens development. *Curr. Mol. Med.* 2010; 10(9): 851-63.
48. Yuan L., Sui T., Chen M., et al. CRISPR/Cas9-mediated GJA8 knockout in rabbits recapitulates human congenital cataracts. *Sci Rep.* 2016; 6: 22024.
49. Zhao D., Wang H.B., Yang J.F., et al. Effect of vascular endothelial growth factor 165 gene transfection on bone defects and its mRNA expression in rabbits. *Chin. Med. J. (Engl)*. 2007; 120(13): 1187-91.
50. Jiang W., Liu L., Chang Q., et al. Production of Wilson disease model rabbits with homology-directed precision point mutations in the ATP7B gene using the CRISPR/Cas9 system. *Sci. Rep.* 2018; 8(1): 1332.
51. Koloskova E.M., Karkischenko V.N., Yezersky V.A., Petrova N.V., Maksimenko S.V., Matveyenko E.L. Rabbit biomodels of human diseases developed using new genomic technologies. *Biomedicine*. 2019; 15(4): 12-33.
52. Yang D., Yang H., Li W., et al. Generation of PPARγ mono-allelic knockout pigs via zinc-finger nucleases and nuclear transfer cloning. *Cell Res. Cell Res.* 2011; 21(6): 979-82.
53. Umeyama K., Watanabe K., Watanabe M., et al. Generation of heterozygous fibrillin-1 mutant cloned pigs from genome-edited foetal fibroblasts. *Sci. Rep.* 2016; 6: 24413.
54. Yan S., Tu Z., Liu Z., et al. A huntingtin knockin pig model recapitulates features of selective neurodegeneration in Huntington's disease. *Cell*. 2018; 173(4): 989-1002.e13.
55. Carlson D.F., Tan W., Lillico S.G., et al. Efficient TALEN-mediated gene knockout in livestock. *Proc. Natl. Acad. Sci. USA*. 2012; 109(43): 17382-7.
56. Huang L., Hua Z., Xiao H., et al. CRISPR/Cas9-mediated ApoE$^{-/-}$and LDLR$^{-/-}$double gene knockout in pigs elevates serum LDL-C and TC levels. *Oncotarget.* 2017; 8(23): 37751-60.
57. Lee B., Dennis J.A., Healy P.J., et al. Hepatocyte gene therapy in a large animal: a neonatal bovine model of citrullinemia. *Proc. Natl. Acad. Sci. USA*. 1999; 96(7): 3981-6.
58. Pinnapureddy A.R., Stayner C., McEwan J., et al. Large animal models of rare genetic disorders: sheep as phenotypically relevant models of human genetic disease. *Orphanet. J. Rare Dis.* 2015; 10: 107.
59. Klebe S., Sykes P.J., Coster D.J., et al. Gene transfer to ovine corneal endothelium. *Clin. Exp. Ophthalmol.* 2001; 29(5): 316-22.

60. Gray-Edwards H.L., Randle A.N., Maitland S.A., et al. Adeno-associated virus gene therapy in a sheep model of Tay–Sachs disease. *Hum. Gene Ther.* 2018; 29(3): 312-26.
61. Howell J.M.C., Walker K.R., Nalbantoglou J., et al. GO 8 Gene therapy trials in the ovine model of McArdle's disease. *Neuromuscul. Disord.* 2006; 16(9): 725.
62. Frisbie D.D., Ghivizzani S.C., Robbins P.D., et al. Treatment of experimental equine osteoarthritis by in vivo delivery of the equine interleukin-1 receptor antagonist gene. *Gene. Ther.* 2002; 9(1): 12-20.
63. Heinzerling L.M., Feige K., Rieder S., et al. Tumor regression induced by intratumoral injection of DNA coding for human interleukin 12 into melanoma metastases in gray horses. *J. Mol. Med. (Berl)*. 2001; 78(12): 692-702.
64. Narfström K., Holland Deckman K., Menotti-Raymond M. The domestic cat as a large animal model for characterization of disease and therapeutic intervention in hereditary retinal blindness. *J. Ophthalmol.* 2011; 2011: 906943.
65. Bradbury A.M., Gurda B.L., Casal M.L., et al. A review of gene therapy in canine and feline models of lysosomal storage disorders. *Hum. Gene Ther. Clin. Dev.* 2015; 26(1): 27-37.
66. McCurdy V.J., Johnson A.K., Gray-Edwards H.L., et al. Sustained normalization of neurological disease after intracranial gene therapy in a feline model. *Sci. Transl. Med.* 2014; 6(231): 231ra48.
67. Hinderer C., Bell P., Gurda B.L., et al. Intrathecal gene therapy corrects CNS pathology in a feline model of mucopolysaccharidosis I. *Mol. Ther.* 2014; 22(12): 2018-27.
68. Hinderer C., Bell P., Gurda B.L., et al. Liver-directed gene therapy corrects cardiovascular lesions in feline mucopolysaccharidosis type I. *Proc. Natl. Acad. Sci. USA*. 2014; 111(41): 14894-9.
69. Foley B.T., Korber B.T.M., Leitner T.K., et al. *HIV Sequence Compendium 2018. № LA-UR-18-25673.* Los Alamos: Los Alamos National Lab. (LANL); 2018.
70. Sundstrom M., Chatterji U., Schaffer L., et al. Feline immunodeficiency virus OrfA alters gene expression of splicing factors and proteasome-ubiquitination proteins. *Virology*. 2008; 371(2): 394-404.
71. Miller C., Abdo Z., Ericsson A., et al. Applications of the FIV model to study HIV pathogenesis. *Viruses*. 2018; 10(4): 206.
72. Giannecchini S., Isola P., Sichi O., et al. AIDS vaccination studies using an ex vivo feline immunodeficiency virus model: failure to protect and possible enhancement of challenge infection by four cell-based vaccines prepared with autologous lymphoblasts. *J. Virol.* 2002; 76(14): 6882-92.
73. Karlas J.A., Siebelink K.H., Peer M.A., et al. Vaccination with experimental feline immunodeficiency virus vaccines, based on autologous infected cells, elicits enhancement of homologous challenge infection. *J. Gen. Virol.* 1999; 80(Pt. 3): 761-5.
74. Ellinwood N.M., Ausseil J., Desmaris N., et al. Safe, efficient, and reproducible gene therapy of the brain in the dog models of Sanfilippo and Hurler syndromes. *Mol. Ther.* 2011; 19(2): 251-9.
75. Nichols B.M., Aoki Y., Kuraoka M., et al. Multi-exon skipping using cocktail antisense oligonucleotides in the canine X-linked muscular dystrophy. *J. Vis. Exp.* 2016; (111): 53776.
76. Connelly S., Mount J., Mauser A., et al. Complete short-term correction of canine hemophilia A by in vivo gene therapy. *Blood.* 1996; 88(10): 3846-53.
77. Morral N., O'Neal W., Zhou H., et al. Immune responses to reporter proteins and high viral dose limit duration of expression with adenoviral vectors: comparison of E2a wild type and E2a deleted vectors. *Hum. Gene Ther.* 1997; 8(10): 1275-86.
78. Gallo-Penn A.M., Shirley P.S., Andrews J.L., et al. Systemic delivery of an adenoviral vector encoding canine factor VIII results in short-term phenotypic correction, inhibitor development, and biphasic liver toxicity in hemophilia A dogs. *Blood.* 2001; 97(1): 107-13.

79. McCormack W.M., Seiler M.P., Bertin T.K., et al. Helper-dependent adenoviral gene therapy mediates long-term correction of the clotting defect in the canine hemophilia A model. *J. Thromb. Haemost.* 2006; 4(6): 1218-25.
80. Marcos-Contreras O.A., Smith S.M., Bellinger D.A., et al. Sustained correction of FVII deficiency in dogs using AAV-mediated expression of zymogen FVII. *Blood.* 2016; 127(5): 565-71.
81. Felsburg P.J., Somberg R.L., Hartnett B.J., et al. Full immunologic reconstitution following nonconditioned bone marrow transplantation for canine X-linked severe combined immunodeficiency. *Blood*. 1997; 90(8): 3214-21.
82. Vernau W., Hartnett B.J., Kennedy D.R., et al. T cell repertoire development in XSCID dogs following nonconditioned allogeneic bone marrow transplantation. *Biol. Blood Marrow Transplant*. 2007; 13(9): 1005-15.
83. Kennedy D.R., Hartnett B.J., Kennedy J.S., et al. Ex vivo γ-retroviral gene therapy of dogs with X-linked severe combined immunodeficiency and the development of a thymic T cell lymphoma. *Vet. Immunol. Immunopathol.* 2011; 142(1-2): 36-48.
84. Ting-De Ravin S.S., Kennedy D.R., Naumann N., et al. Correction of canine X-linked severe combined immunodeficiency by in vivo retroviral gene therapy. *Blood*. 2006; 107(8): 3091-7.
85. Felsburg P.J., De Ravin S.S., Malech H.L., et al. Gene therapy studies in a canine model of X-linked severe combined immunodeficiency. *Hum. Gene Ther. Clin. Dev.* 2015; 26(1): 50-6.
86. Burtner C.R., Beard B.C., Kennedy D.R., et al. Intravenous injection of a foamy virus vector to correct canine SCID-X1. *Blood*. 2014; 123(23): 3578-84.
87. Acland G.M., Aguirre G.D., Ray J., et al. Gene therapy restores vision in a canine model of childhood blindness. *Nat. Genet.* 2001; 28(1): 92-5.
88. Bianco S.R., Sun J., Fosmire S.P., et al. Enhancing antimelanoma immune responses through apoptosis. *Cancer Gene Ther.* 2003; 10(9): 726-36.
89. Chen Y., Niu Y., Ji W. Genome editing in nonhuman primates: approach to generating human disease models. *J. Int. Med.* 2016; 280(3): 246-51.
90. Chen Y., Yu J., Niu Y., et al. Modeling Rett syndrome using TALEN-edited MECP2 mutant cynomolgus monkeys. *Cell*. 2017; 169(5): 945-955.e10.
91. Chen Y., Zheng Y., Kang Y., et al. Functional disruption of the dystrophin gene in rhesus monkey using CRISPR/Cas9. *Hum. Mol. Genet.* 2015; 24(13): 3764-74.
92. Ott de Bruin L.M., Volpi S., Musunuru K. Novel genome-editing tools to model and correct primary immunodeficiencies. *Front. Immunol.* 2015; 6: 250.
93. Sato K., Oiwa R., Kumita W., et al. Generation of a nonhuman primate model of severe combined immunodeficiency using highly efficient genome editing. *Cell Stem Cell.* 2016; 19(1): 127-38.
94. Gheit T., Sekkat S., Cova L., et al. Experimental transfection of Macaca sylvanus with cloned human hepatitis B virus. *J. Gen. Virol.* 2002; 83(Pt. 7): 1645-9.
95. Yang C., Ruan P., Ou C., et al. Chronic hepatitis B virus infection and occurrence of hepatocellular carcinoma in tree shrews (Tupaia belangeri chinensis). *Virol. J.* 2015; 12: 26.
96. Peters A., Rosene D.L., Moss M.B., et al. Neurobiological bases of age-related cognitive decline in the rhesus monkey. *J. Neuropathol. Exp. Neurol.* 1996; 55(8): 861-74.
97. Price D.L., Martin L.J., Sisodia S.S., et al. Aged non-human primates: an animal model of age-associated neurodegenerative disease. *Brain Pathol.* 1991; 1(4): 287-96.
98. Zhao Q., Lu J., Yao Z., et al. Upregulation of Aβ42 in the brain and bodily fluids of rhesus monkeys with aging. *J. Mol. Neurosci.* 2017; 61(1): 79-87.
99. Geula C., Nagykery N., Wu C.K. Amyloid-β deposits in the cerebral cortex of the aged common marmoset (Callithrix jacchus): incidence and chemical composition. *Acta Neuropathol.* 2002; 103(1): 48-58.

100. Voytko M.L., Olton D.S., Richardson R.T., et al. Basal forebrain lesions in monkeys disrupt attention but not learning and memory. *J. Neurosci.* 1994; 14(1): 167-86.
101. Teng E., Stefanacci L., Squire L.R., et al. Contrasting effects on discrimination learning after hippocampal lesions and conjoint hippocampal–caudate lesions in monkeys. *J. Neurosci.* 2000; 20(10): 3853-63.
102. Tuszynski M.H., Roberts J., Senut M.C., et al. Gene therapy in the adult primate brain: intraparenchymal grafts of cells genetically modified to produce nerve growth factor prevent cholinergic neuronal degeneration. *Gene Ther.* 1996; 3(4): 305-14.
103. Tuszynski M.H., Yang J.H., Barba D., et al. Nerve growth factor gene therapy: activation of neuronal responses in Alzheimer disease. *JAMA Neurol.* 2015; 72(10): 1139-47.
104. Yang S.H., Cheng P.H., Banta H., et al. Towards a transgenic model of Huntington's disease in a non-human primate. *Nature*. 2008; 453(7197): 921-4.
105. Lasbleiz C., Mestre-Francés N., Devau G., et al. Combining gene transfer and nonhuman primates to better understand and treat Parkinson's disease. *Front. Mol. Neurosci.* 2019; 12: 10.
106. Przedbroski S., Levivier M., Jiang H., et al. Dose-dependent lesions of the dopaminergic nigrostriatal pathway induced by instrastriatal injection of 6-hydroxydopamine. *Neuroscience*. 1995; 67(3): 631-47.
107. Schober A., Classic toxin-induced animal models of Parkinson's disease: 6-OHDA and MPTP. *Cell Tissue Res.* 2004; 318(1): 215-24.
108. Petroske E., Meredith G.E., Callen S., et al. Mouse model of Parkinsonism: a comparison between subacute MPTP and chronic MPTP/probenecid treatment. *Neuroscience*. 2001; 106(3): 589-601.
109. Deng H., Yuan L. Genetic variants and animal models in SNCA and Parkinson disease. *Ageing Res. Rev.* 2014; 15: 161-76.
110. Ramonet D., Daher J.P., Lin B.M., et al. Dopaminergic neuronal loss, reduced neurite complexity and autophagic abnormalities in transgenic mice expressing G2019S mutant LRRK2. *PLoS One.* 2011; 6(4): e18568.
111. Blesa J., Przedborski S. Parkinson's disease: animal models and dopaminergic cell vulnerability. *Front. Neuroanat*. 2014; 8: 155.
112. Kirik D., Annett L.E., Burger C., et al. Nigrostriatal α-synucleinopathy induced by viral vector-mediated overexpression of human α-synuclein: a new primate model of Parkinson's disease. *Proc. Natl. Acad. Sci. USA*. 2003; 100(5): 2884-9.
113. Koprich J.B., Johnston T.H., Reyes G., et al. Towards a non-human primate model of alpha-synucleinopathy for development of therapeutics for Parkinson's disease: optimization of AAV1/2 delivery parameters to drive sustained expression of alpha synuclein and dopaminergic degeneration in macaque. *PLoS One.* 2016; 11(11): e0167235.
114. Yang W., Wang G., Wang C.E., et al. Mutant alpha-synuclein causes age-dependent neuropathology in monkey brain. *J. Neurosci.* 2015; 35(21): 8345-58.
115. Junyent F., Kremer E.J. CAV-2 — why a canine virus is a neurobiologist's best friend. *Curr. Opin. Pharmacol.* 2015; 24: 86-93.
116. Mestre-Francés N., Serratrice N., Gennetier A., et al. Exogenous LRRK2G2019S induces parkinsonian-like pathology in a nonhuman primate. *JCI Insight.* 2018; 3(14): e98202.
117. McBride J.L., Neuringer M., Ferguson B., et al. Discovery of a CLN7 model of Batten disease in non-human primates. *Neurobiol. Dis.* 2018; 119: 65-78.

Conclusion

Targeted genome editing with programmable nucleases has moved to the frontline of genetic technology and is widely used in various fields related to genome modification. The discoveries of recent years have shown that targeted genome editing systems can carry enormous potential that can be applied to modern science and medicine. Targeted genome editing can be used for developing cell models of hereditary diseases in humans and animals, for functional screening of genomes, for researching epigenomes and visualization of cellular processes as well as in the food industry for producing healthy food products and in agriculture for developing livestock breeds and plant varieties. Targeted genome editing systems are instrumental in developing approaches for treatment of genetic, infectious, oncological, immunological, and other diseases that were previously considered incurable.

Targeted genome editing technologies can be used as antiviral therapy and in the fight against persistent viral infections caused by HIV, hepatitis B virus, Epstein-Barr virus, cytomegalovirus, human papillomaviruses, *etc*. In addition, genomic editing allows correcting mutations that are a direct cause of a particular disease. Approaches have been developed for the treatment of Duchenne muscular dystrophy, epidermolysis bullosa, sickle cell disease, beta-thalassemia, xeroderma pigmentosa, and other monogenic human diseases. In addition, targeted genome editing systems represent a revolutionary tool for the treatment of malignant neoplasms, with enormous potential for the development of new gene and cell therapy methods.

Clinical trials of candidate therapeutic products are being conducted worldwide; these products are based on programmable nucleases — the main components of targeted genome editing systems. The efficiency of gene and cell therapy is being tested in combating hematological and solid neoplasms, hereditary and infectious diseases. There is no doubt that the number of clinical trials evaluating therapeutic products developed by using targeted genome editing technologies will be steadily increasing.

Some of the targeted genome editing systems can be used to diagnose diseases by identifying genetic sequences, such as viruses or oncogenes, and to prevent infectious diseases by modifying genes in disease carriers, for example, malaria transmitters, rather than in humans. In future, diagnostic systems based on targeted genome editing will be used not only for the qualitative and/or quantitative assessment of nucleic acids instead of specific PCR analyzes, but also for rapid multiplex assessment of RNA expression, detection of sample contamination with nucleic acids, and tracking of transcripts associated with pathological processes within biological systems. Such diagnostic systems can provide multifunctional, error-tolerant methods for prompt diagnosis, including infectious diseases and sensitive genotyping of microorganisms, at the bedside and in the field.

Many impressive breakthroughs have been made since the discovery and description of the first systems ushering in targeted genome editing. Every day, methods of targeted genome editing are improving and developing. Technological progress and science advancement improve the safety and the efficiency of programmable nucleases, techniques developed for delivery elements of genome editing systems, and methods used for identification of on-target and off-target effects of programmable nucleases. Targeted genome editing systems have become an integral part of the scientific progress, and scientists keep working on looking for and discovering new gene-editing tools.

РОССИЙСКАЯ ФЕДЕРАЦИЯ

ПАТЕНТ

НА ИЗОБРЕТЕНИЕ

№ 2707542

Способ получения препарата рекомбинантной нуклеазы CAS, по существу, свободного от бактериальных эндотоксинов, полученный данным способом препарат и содержащий его набор для использования в системе CRISPR/Cas

Патентообладатель: ***Федеральное бюджетное учреждение науки "Центральный научно-исследовательский институт эпидемиологии" Федеральной службы по надзору в сфере защиты прав потребителей и благополучия человека (ФБУН ЦНИИ Эпидемиологии Роспотребнадзора) (RU)***

Авторы: ***см. на обороте***

Заявка № **2019109018**
Приоритет изобретения **28 марта 2019 г.**
Дата государственной регистрации в Государственном реестре изобретений Российской Федерации **27 ноября 2019 г.**
Срок действия исключительного права на изобретение истекает **28 марта 2039 г.**

Руководитель Федеральной службы по интеллектуальной собственности

Г.П. Ивлиев

РОССИЙСКАЯ ФЕДЕРАЦИЯ

ПАТЕНТ

НА ИЗОБРЕТЕНИЕ

№ 2720767

Способ обнаружения провирусной ДНК вируса иммунодефицита человека, интегрированной в геном человека, в ультранизких концентрациях и специфические олигонуклеотиды для использования в способе

Патентообладатель: ***Федеральное бюджетное учреждение науки «Центральный научно-исследовательский институт эпидемиологии» Федеральной службы по надзору в сфере защиты прав потребителей и благополучия человека (ФБУН ЦНИИ Эпидемиологии Роспотребнадзора) (RU)***

Авторы: ***см. на обороте***

Заявка № **2019141298**
Приоритет изобретения **13 декабря 2019 г.**
Дата государственной регистрации в Государственном реестре изобретений Российской Федерации **13 мая 2020 г.**
Срок действия исключительного права на изобретение истекает **13 декабря 2039 г.**

Руководитель Федеральной службы по интеллектуальной собственности

Г.П. Ивлиев

РОССИЙСКАЯ ФЕДЕРАЦИЯ

ПАТЕНТ

НА ИЗОБРЕТЕНИЕ

№ 2720768

Система CRISPR-Cas для детекции провирусной ДНК вируса иммунодефицита человека, интегрированной в геном человека, в ультранизких концентрациях

Патентообладатель: ***Федеральное бюджетное учреждение науки "Центральный научно-исследовательский институт эпидемиологии" Федеральной службы по надзору в сфере защиты прав потребителей и благополучия человека (ФБУН ЦНИИ Эпидемиологии Роспотребнадзора) (RU)***

Авторы: ***см. на обороте***

Заявка № **2019141293**
Приоритет изобретения **13 декабря 2019 г.**
Дата государственной регистрации в Государственном реестре изобретений Российской Федерации **13 мая 2020 г.**
Срок действия исключительного права на изобретение истекает **13 декабря 2039 г.**

Руководитель Федеральной службы по интеллектуальной собственности

Г.П. Ивлиев

РОССИЙСКАЯ ФЕДЕРАЦИЯ

ПАТЕНТ

НА ИЗОБРЕТЕНИЕ

№ 2720769

Способ получения препарата рибонуклеопротеинового комплекса CRISPR/CAS и препарат для детекции провирусной ДНК вируса иммунодефицита человека, интегрированной в геном человека, в ультранизких концентрациях

Патентообладатель: ***Федеральное бюджетное учреждение науки «Центральный научно-исследовательский институт эпидемиологии» Федеральной службы по надзору в сфере защиты прав потребителей и благополучия человека (ФБУН ЦНИИ Эпидемиологии Роспотребнадзора) (RU)***

Авторы: ***см. на обороте***

Заявка № **2019141306**
Приоритет изобретения **13 декабря 2019 г.**
Дата государственной регистрации в Государственном реестре изобретений Российской Федерации **13 мая 2020 г.**
Срок действия исключительного права на изобретение истекает **13 декабря 2039 г.**

Руководитель Федеральной службы по интеллектуальной собственности

Г.П. Ивлиев

GENETIC TECHNOLOGIES

Edited by
Vasily G. Akimkin
Full Member of the Russian Academy of Sciences

Translated by Retneva-Schmaisser O.N., individual entrepreneur
Translation editing by Online perevody, Ltd.
Publishing editor Olga V. Osokina
Russian editor Olga V. Ustinkova
Illustrator Natalia Yu. Palochkina
Layout Vadim I. Arkhipov
Proofreading Yulia V. Smirnova
Bibliographer Darya G. Filatova

Central Research Institute for Epidemiology of Rospotrebnadzor
111123, 3A, Novogireevskaya str., Moscow, Russian Federation. www.crie.ru

Printed in Ob'yedinennyy poligraficheskiy kompleks, Ltd.
www.opk.bz. E-mail: info@opk.bz
Printed 300 copies.