European Commission

JRC TECHNICAL REPORT

New Genomic Techniques: State-of-the-Art Review

Broothaerts, W., Jacchia, S., Angers, A., Petrillo, M., Querci, M., Savini, C., Van den Eede, G. and Emons, H.

2021

Joint Research Centre

EUR 30430 EN

Contact information
European Commission, Joint Research Centre
Retieseweg 111, 2440 Geel, Belgium
Email: Wim.Broothaerts@ec.europa.eu

EU Science Hub
https://ec.europa.eu/jrc

JRC121847

EUR 30430 EN

PDF ISBN 978-92-76-24696-1 ISSN 1831-9424 doi:10.2760/710056

Luxembourg: Publications Office of the European Union, 2021

How to cite this report: Broothaerts, W., Jacchia, S., Angers, A., Petrillo, M., Querci, M., Savini, C., Van den Eede, G. and Emons, H., *New Genomic Techniques: State-of-the-Art Review*, EUR 30430 EN, Publications Office of the European Union, Luxembourg, 2021, ISBN 978-92-76-24696-1, doi:10.2760/710056, JRC121847.

Contents

Authors

Wim Broothaerts, Sara Jacchia, Alexandre Angers, Mauro Petrillo, Maddalena Querci, Cristian Savini, Guy Van den Eede & Hendrik Emons

Abstract

This review of the scientific and technological developments on New Genomic Techniques (NGTs) is aimed to provide the technical status of NGTs with respect to their diverse mechanisms of action and applicability. It has been compiled in support to the request to the Commission to submit a study in light of the Court of Justice's judgment in Case C-528/16 regarding the status of novel genomic techniques under Union law (Council Decision (EU) 2019/1904).

Since the adoption of Directive 2001/18/EC a variety of NGTs has been developed which are capable to alter the genome of an organism. These techniques aim to modify plant, animal or microbial organisms by accelerating the breeding or development process, and/or by rendering it more precise. They are seen as promising instruments for the agri-food and industrial biotechnology sectors, but they are also offering tremendous innovative potential and technical possibilities for the health sector.

This study used a systematic literature survey to identify the major NGTs employed for genome modifications in plants, animals and microorganisms.

Many of the NGTs are built on the versatile CRISPR-Cas technology, which can be used in different versions and to which additional functionalities may be added.

NGTs may affect only single nucleotide changes or may delete, replace or insert very large sequences and thus a classification of NGTs on the basis of the size of the nuclear fragment affected is not feasible.

Here, we have therefore developed a classification system by arranging the NGTs into four groups based on the interaction of their active components with the genome.

Furthermore, we give an overview of the possible genome alterations and their likelihood of occurring in nature or through conventional breeding. We also provide the main NGTs that may be involved to generate those genome alterations in the different kingdoms of living organisms (bacteria, fungi, plants and animals). It has to be noted that NGTs are continuously being modified for an improved performance and the field is still evolving in a very dynamic manner. Therefore, this review is providing a non-exhaustive list of NGTs without any implicit legal judgement on their status under current EU legislation.

Executive summary

This study provides the current state-of-the-art knowledge on scientific developments in the field of new genomic techniques (NGTs). NGTs are here defined as 'techniques that are able to alter the genetic material of an organism, developed after the publication of EU Directive 2001/18/EC'. The study, requested from other Commission services to the Joint Research Centre, is representing one element of the broader study that the European Commission is preparing on this topic upon request of the Council of the European Union. This review provides a non-exhaustive list of NGTs without any implicit legal judgement on their status under current Union law.

All living organisms harbour a genome composed of nucleotide sequences in the form of DNA or RNA. The genome of a species is not characterised by a fixed nucleotide sequence but rather by a sequence scaffold with millions of sequence variations appearing in different individuals belonging to the same species. NGTs can add sequence variations to the genome of an organism (a process called 'genome editing'). Such variations may be entirely novel and may or may not occur already in other individuals of the species. NGTs may also be used to introduce new sequences derived from other species into an organism.

Scientific advances in molecular biology during the past 20 years have deciphered the molecular mechanisms of many functional properties in various organisms and their genetic basis. Together with improved next-generation sequencing (NGS) technologies, the technical basis is available for the manipulation of an organism's genome. Whereas several established techniques of genomic modification generate random sequence alterations in the genome, new technological developments can make changes directed to a selected genomic location and thus allow a more precise editing of the genome. This genome editing has rapidly revolutionised the fields of breeding in plants, animals and molecular engineering of microorganisms in recent years and provides also new opportunities for gene therapy in humans.

This study reviews the major NGTs and describes their characteristics. The identification of the various NGTs is based on a systematic literature review that was concluded on 31 January 2020. For the detailed description of each NGT additional literature was retrieved through targeted searches between March and June 2020.

Four groups of NGTs are distinguished based on interactions with the genome:

1) NGTs creating a double-strand break in the DNA;
2) NGTs achieving genome editing without breaking the DNA double helix or generating only a single-strand DNA break;
3) NGTs inducing epigenomic changes;
4) NGTs acting specifically on RNA.

Within each group several NGTs are described with their mode of action, induced modifications and the organisms to which the technique can be applied. Information on possible unintended (usually called 'off-target') modifications and limitations in our current understanding complement the technical descriptions for each NGT.

A NGT may generate alterations depending on the particular way the technique is employed. On the other hand, the same alteration may be achieved by applying different NGTs, but not all of them work efficiently in all types of organisms. Furthermore, not every desired alteration can be readily achieved at any sequence because some NGTs may have restrictions for the recognition and binding of their targets.

Some techniques, like those based on the use of site-directed nucleases (SDNs), only induce a double-strand break (DSB) at a selected site in the genome. The actual genome editing is then a result of the repair of these DSBs by one of several endogenous cellular mechanisms, which occasionally create mutations. These techniques can be used with or without an added donor sequence that may function as template during the repair processes. The integration in the genome of a heterologous sequence used as donor template may result from homologous recombination but also from other strand break repair pathways.

Other techniques use either catalytically impaired SDNs that generate only a single-strand break in the DNA or SDNs with completely abolished cleavage activity that only recognise and bind a target sequence. Coupling the impaired SDNs with another enzyme allows single base substitutions, deletions and sequence replacements at chosen sites in DNA, methylation changes of the DNA or histone modifications, or RNA editing, cleavage or alternative splicing. Other techniques involve oligonucleotides for DNA editing or RNA interference, or engineered recombinase-mediated systems exploited from bacteriophages.

The most prominent set of NGTs is based on the CRISPR-Cas site-directed nuclease technology that exponentially expanded the opportunities for the modification of many genomic targets in diverse organisms. It is a versatile and relatively easy to implement technology for genome editing, which can also be used for simultaneous editing at multiple sites. It has been used as a platform for many of the other NGTs by adding different functionalities to it. Not all of these derived NGTs have been extensively studied in different experimental settings and their utility may not be completely understood at the present time. The CRISPR-Cas based techniques are still evolving and the list of NGTs is expected to further expand in the coming years.

For the NGT to be functional, its active components, if not already present, need to be delivered to the cell to be treated. Some NGTs function upon the sole provision of oligonucleotides to create small DNA or RNA changes. Other NGTs may require delivery of transgenes by recombinant DNA technology involving biolistic, bacterial or viral delivery systems, leading to transient vector expression or stable transgene integration. An alternative to the use of DNA vectors is the temporary delivery of mRNA or RNA-protein complexes to the cells, providing, as a consequence, temporary NGT functionalities.

Whereas the NGTs described herein aim for alterations at specific target sites, off-target alterations have been reported in the literature when NGTs were used. Various bioinformatics tools have been developed to screen for potential off-target sites in a particular genome and to predict the probability of their occurrence. In the animal field, and particularly for human therapy applications, various efforts have been successfully undertaken to improve the target efficiency of various NGTs and to reduce the occurrence of off-target alterations. Many of these studies have been carried out on cultured cells. Target specificity has received less emphasis in plants because unwanted off-target alterations could be segregated in subsequent sexual crossings.

This study highlights that many NGTs may be used in different forms, *e.g.* with functional or catalytically impaired enzymes, with or without donor template, targeting one or multiple sites and using different delivery systems, with distinct outcomes in terms of the type and size of the resulting genome alterations. Next-generation NGTs and further improvements to current NGTs are to be expected in the coming years. This will likely increase the efficiency and specificity of genome editing in various organisms and will expand the opportunities for agricultural breeding, industrial biotechnology and human gene therapies.

1 Introduction

Genetic variations in nature are the result of changes ('mutations') in the genetic material of an organism. Whereas most of these changes may have no direct effect or are deleterious, some of them allow organisms to adapt to changing environmental conditions and drive evolution through natural selection.

Since the origin of agriculture, people have exploited the genetic variability existing in nature and have developed abilities to select, combine and/or introduce new genetic variations.

It was only after Gregor Mendel outlined the principles of heredity that the breeding processes became science-based. When in the early 20th century the principles of genetic inheritance were further understood, a beginning was made towards applying them for the improvement of plants and animals and since then many new scientific discoveries in heredity gave rise to new types of breeding applications.

It was found that the chances for an abortion of seed embryos increased the greater the distance between species crossed. Therefore, an *in vitro* technique called 'embryo rescue', developed around 1925 to rescue inherently weak, immature or hybrid embryos to prevent degeneration, was a milestone for modern breeding. The technique has been used, for instance, to breed for seedless grapes (reviewed in Li *et al.*, 2015).

One idea developed by Stadler in 1928 (Stadler, 1928) was to induce random mutations in crop plants and to observe the effects. By the 1950s the so-called atomic gardening, where plant seeds were exposed to irradiation, became a very popular application of fission energy (Howorth, 1960).

A first borderline was passed with such atomic gardening: breeders did no longer select and combine naturally occurring mutations, but they started inducing them randomly with physical and chemical means. Unlike with naturally occurring mutations, which result in one or a few mutations per generation, these techniques can induce very high numbers (often many thousands) of simultaneous mutations. Therefore, crops that did not previously occur in nature could be created (Ahloowalia *et al.*, 2004). A second borderline was passed in the early seventies, when genetic material from one organism was successfully integrated into another – unrelated – organism (Cohen *et al.*, 1973). Indeed, 'breeding' concerned so far only the mixing of genes within one species, but in principle 'recombinant DNA' may combine the genetic material from any given species and thus it crosses the natural genetic barriers.

These two borderlines form the basis of the current EU legislation on GMOs, where a 'genetically modified organism (GMO)' means "*an organism, with the exception of human beings, in which the genetic material has been altered in a way that does not occur naturally by mating and/or natural recombination*"[1].

Scientific discoveries in molecular biology have continued at a fast pace and a variety of advanced techniques have been developed in the last 20 years, boosted by the scientific progress in DNA sequencing technology and bioinformatic analysis tools (Koboldt *et al.*, 2013). Long-read sequencing technology has resulted in the discovery of hundreds of thousands of structural variants (long insertions, deletions, duplications and chromosomal rearrangements) between breeding lines and of their relationship with agriculturally important traits (Alonge *et al.*, 2020).

In addition, understanding the molecular basis of certain traits provides opportunities for the creation of new genetic variations in breeding programmes by copying the genetic information found in organisms displaying interesting traits to the genome of an elite cultivar or breed. Similarly, malfunctioning genes in plants, animals and humans may be cured, if the genetic basis of the disease phenotype has been elucidated.

Manipulating the genetic makeup of an organism has now become increasingly precise and new genomic techniques can introduce genetic alterations ranging from single nucleotide changes to large sequence deletions and insertions. The developing biotechnology landscape contains a growing number of such molecular techniques.

In 2011, the Joint Research Centre (JRC) of the European Commission reviewed upon request of DG SANTE the state-of-the-art of some of the emerging new plant breeding techniques, their level of development and adoption by the breeding sector and the prospects for a future commercialisation of plants created by them (Lusser *et al.*, 2011 & 2012).

[1] Directive 2001/18/EC of the European Parliament and of the Council of 12 March 2001 on the deliberate release into the environment of genetically modified organisms and repealing Council Directive 90/220/EC, Off. J. L 106, 17.04.2001, p. 1-39.

In March 2019, the European Network of GMO Laboratories (ENGL), coordinated by the JRC, published a report on the challenges for the detection of food and feed plant products obtained by new mutagenesis techniques, which was established in cooperation with the EU Reference Laboratory for GM Food and Feed (ENGL, 2019).

On 8 November 2019 the Council of the EU adopted a Decision requesting the Commission to submit "a study by 30 April 2021 regarding the status of new genomic techniques under Union Law and in light of the Court of Justice's judgment of 25 July 2018[2]. The Council also requested the Commission to submit "a proposal (accompanied by an impact assessment), if appropriate in view of the outcomes of the study, or otherwise to inform the Council on other measures required as a follow-up of the study". As a contribution to this study, DG SANTE requested the JRC to provide a study on the current and future scientific developments and market applications of new genomic techniques.

This review describes the scientific and technological developments of **New Genomic Techniques** (NGTs), which are here defined as '*techniques which are capable to alter the genetic material of an organism, and which have been mainly developed after the publication of Directive 2001/18/EC*[1]'. Genomic techniques developed prior to the publication of this Directive are called here **Established Genomic Techniques** (EGTs).

In the following, the different types of NGTs are outlined, including how they differ from EGTs, the diverse outcome of the different techniques and how this relates to the natural genetic variation in organisms' genomes and the way the NGT components are delivered to the cells. For each of the NGTs, the mode of action, the alterations generated and the observation of off-target alterations, the type of organisms to which the technique has been applied and how it is delivered are described in detail.

The study has exploited a systematic literature search methodology to define and classify the techniques, supplemented with in-depth targeted literature searches on the internet. The scope of this study covers scientific developments in NGTs for the agri-food, pharmaceutical and industrial biotechnology sectors including plants, animals and microorganisms.

This document constitutes one of two JRC studies and reviews the current scientific and technical status of technologies and approaches used. A second study addresses the market applications of NGTs These JRC studies will form part of the larger study that the European Commission is preparing in response to the Council's request.

[2] The Court of Justice of the EU ruled in Case C-528/16 that new mutagenesis techniques fall within the scope of Directive 2001/18/EC and are subject to the obligations laid down therein (ECLI:EU:C:2018:583).

2 Terminology and abbreviations

2.1 Definitions used in this document

Here we define genomic techniques as 'techniques which are capable to alter the genetic material of an organism' and we further distinguish:

- Established Genomic Techniques (EGTs): genomic techniques developed prior to the publication of Directive 2001/18/EC (April 2001)
- New Genomic Techniques (NGTs): genomic techniques developed after the publication of Directive 2001/18/EC (April 2001)

It must be emphasised that 'technique' and 'technology' encompass different concepts. A technology (*e.g.* recombinant DNA technology or next-generation sequencing technology) refers to the application of scientific knowledge for practical purposes, whereas a technique refers to a particular method or methodology (*e.g. Agrobacterium*-mediated plant transformation is an established genomic technique).

2.2 Abbreviations

AAV	adeno-associated virus
ABE	adenine base editor
ADAR	adenosine deaminase that act on RNA enzymes
ASO	antisense oligonucleotide
BE	base editing or base editor
BIR	break induced repair
bp	basepair
Cas	CRISPR associated system
CBE	cytosine base editor
CRISPR	clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
dCas	dead (or catalytically deactivated) Cas (due to mutation in both nuclease domains)
DNMT	DNA methyltransferase
DSB	double-strand break
dsDNA (dsRNA)	double stranded DNA (or RNA)
HDR	homology-directed repair
HE	homing endonuclease (= meganuclease)
HITI	homology-independent template integration
HR	homologous recombination
kb	kilo basepairs (*i.e.* thousand bp)
indel	insertions and deletions (common abbreviation)
MMEJ	micro-homology mediated end joining
MMR	mismatch repair
nCas	nickase Cas (due to mutation in one of the nuclease domains)
NHEJ	non-homologous end joining
NLS	nuclear localisation signal
nt	nucleotide

ODM	oligonucleotide-directed mutagenesis
PAM	protospacer adjacent motif
PAMmer	antisense oligonucleotide that following annealing to target ssDNA constitutes a PAM site for sgRNA recognition
PE	prime editing
pegRNA	prime editing extended guide RNA
PFS	protospacer flanking site
PITCh	precise integration into target chromosome
RMCE	recombinase-mediated cassette exchange
RNP	ribonucleoprotein
SDN	site-directed nuclease
SDRE	site-directed RNA (base) editing
SDSA	synthesis-dependent strand annealing
sgRNA	single-guide RNA
SSA	single-strand annealing
ssDNA (ssRNA)	single-strand DNA (or RNA)
SSR	site-specific recombination (or recombinase)
SST	site-specific transposition (or transposase)
TAL(E)(N)	transcription activator-like (effector) (nuclease)
TET	ten-eleven translocation demethylase
TFO	triplex-forming oligonucleotide
TIR	terminal inverted repeat
tracRNA	trans-activating CRISPR RNA
ZF(N)	zinc finger (nuclease)

3 Methodology

3.1 Literature search strategy

A literature search was carried out to generate the knowledge basis for the state-of-the-art review of NGTs. This provided an overview of the type of NGTs to be reviewed in this study. Additional scientific papers were identified and retrieved from the web during the writing process.

A flowchart, included as Annex 1, displays the different steps of the search strategy. More details are provided below.

3.1.1 Definition of a query to be used for literature searches (QUERY)

The query that has been designed and used is the following:

QUERY = TITLE-ABS-KEY ((nucleic* OR genom* OR gene*) W/5 (edit* OR manipul* OR engineer* OR modif*) W/5 (technique* OR technolog*))

where "W/5" means that the words have to be at a maximum of 5 words of distance.

3.1.2 Selection of repositories for literature search

We used the QUERY to scan two literature repositories:

- **SCOPUS[3]:** Elsevier's Scopus is the largest abstract and citation database of peer-reviewed literature.
- **EuropePMC[4]:** Europe PubMed Central (EuropePMC) is an open science platform that enables access to a worldwide collection of life science publications and preprints from trusted sources around the globe. It is a partner of PubMed Central and has over 5 million more abstracts than PubMed, as it includes also patents, Agricola records and preprints. In addition, where available, the whole full text is indexed.

These two repositories provide also Application Program Interfaces (APIs) for automatic fetching and retrieval of the metadata associated to each selected document.

The search was carried out on 31 January 2020.

3.1.3 Fetch of information of the identified documents (GROUP-T0)

It was decided to focus on reviews only for the period 2001-2016 and on every retrieved document (original research paper, review or editorial) for the period 2017-today. Only documents in English have been taken into account for this study.

After merging the two sets of identified documents (each from a repository) into a non-redundant list, the final set of GROUP-T0 consisted of 7,397 documents.

3.1.4 Screening of GROUP-T0 set

Documents of Group-T0 were divided into 3 groups and each group was screened by a different JRC scientist on the basis of their titles, in order to create a set of documents to be further inspected (GROUP-T1). The process ended with the creation of a GROUP-T1 set of 1,363 documents.

3.1.5 Screening of GROUP-T1 set

Documents of GROUP-T1 were divided into 3 groups and each group was screened by a different JRC scientist on the basis of the abstracts (making sure each scientist screened a set of articles different from those screened for GROUP-T0), in order to create a final set of 264 documents to be retrieved and read.

[3] https://www.scopus.com/

[4] https://europepmc.org/

3.1.6 Establishment of the CORE set

To the selected 264 papers another 31 key papers were added from our own knowledge base. These additional papers constituted either articles published after the search was done (including pre-print and online publications), *i.e.* from 01 February up to 15 April 2020, or were published in recent, specialised peer-reviewed journals not covered by the databases used, *e.g.* in The CRISPR Journal, or were missed by the search terms used. The final list of 299 papers constituted the CORE set.

3.1.7 Verification of the process

The CORE list of articles was examined for the presence of a number of key articles on NGTs that had been independently collected before this exercise while following this topic, and which were considered relevant publications for the description of the state-of-the-art of NGTs. This allowed to verify whether the process used was in line with the expectations. It was concluded that the majority of the key articles were present in the CORE set and that the key NGTs described in these papers were represented.

3.1.8 Back search from the CORE set

The CORE set was used to trace back the documents referenced in the CORE set in order to generate a second round of fetched documents to be specifically used to describe the state-of-art of each NGT, including known limitations and knowledge gaps. This formed the POSTCORE-1 set, composed of 5,885 unique papers. References for the study have been retrieved by using classification and clustering strategies.

A further round of literature retrieval was done by collecting the references cited in the POSTCORE-1 set. As this POSTCORE-2 list contained 82,395 articles, mostly published >10 years ago and dealing with a broad set of topics, it was decided not to continue with the POSTCORE-2 set.

3.1.9 Selection of articles for the review

The CORE and POSTCORE-1 papers were combined (6,184 papers in total) and a further selection was made to trim down the number of papers to a reasonable set to work with. POSTCORE-1 papers that were published in 2019 and 2020 were retained. For older articles in the POSTCORE-1, only those cited by at least 4 articles in the CORE list were also kept. These were combined to all articles in the CORE list, with duplicates removed.

The selected list of 745 papers was divided into three groups and each group was reviewed by a different JRC scientist on the basis of their title and abstract (and where needed full text). In this process the papers were classified by tentative type of NGT.

During this review, a total of 163 papers could not be directly classified and needed additional detailed reading. These papers were dealing with less common techniques or applications in specific organisms. This set was subdivided into six groups and each group was reviewed in detail by one JRC scientist, checking also the full text of the articles. This additional review clarified if these papers described a NGT, what its characteristics were and why it differed from the NGTs already identified. This information was used to further investigate these NGTs on the basis of the available information. A total of 94 papers out of the 164 papers were discarded.

The final set of papers consisted of 651 papers that were found relevant for the study. This list can be made available upon request.

3.2 Drafting approach

The final set of papers constituted the basis for the identification of the different NGTs and for summarising their characteristics. Additional publications for compiling the information described in this study were obtained by targeted searches during the drafting process (March-June 2020). The report was thoroughly reviewed within the JRC and independently by other Commission services.

Each of the techniques is described in a common way:

- Introduction: introduces the NGT and provides the broader context;
- Mechanism: describes the mechanism of the NGT in general terms, with more details on the mode of action provided in a separate text box for the expert reader;

- Inducible modifications: describes the alterations of the genetic material that can be induced by the NGT;
- Result: describes the result of the induced modifications;
- Target organisms: provides examples of the type of organisms to which the NGT can be applied, mainly based on what has been reported in the literature up to now. This does not mean that the NGT cannot be applied to any other organism in the future;
- Potential off-target effects: reviews the reported off-target alterations generated by the NGT;
- Limitations and gaps in knowledge: reviews any knowledge gaps or requirements for further investigations and the technical limitations of the NGT.

4 New genomic techniques: classification and general characteristics

4.1 Established genomic techniques

Established genomic techniques may be applied *in planta* or *in vitro*, and may result in genetic alterations that might also occur naturally or not.

One set of established genomic techniques are breeding techniques that are employed to hybridise parent plants, which can also cross in nature (sexual hybridisation techniques). To obtain progeny from the desired male and female parentage, it is necessary to master techniques for manipulating the hybridisation process (*e.g.* to cross-pollinate a self-pollinating plant) and some typical interventions are emasculation of flowers, isolation of female flowers and artificial application of viable pollen. Breeding techniques may also be used to overcome spatial barriers (worldwide transport of seed and pollen) and chronological barriers (climatological conditions may be applied to ensure that two races flower at the same time). The resulting genetic alterations can also occur in nature.

Techniques such as ovule culture and *in vitro* fertilisation by electrofusion have been developed to overcome physiological barriers. Such techniques may thus generate hybrids among cross-sterile plants that normally do not occur in nature.

Another sophisticated breeding technique involves polyploidy induction (polyploidy breeding technique) which is the alteration in the number of chromosomes of an organism, either through autopolyploidy, *i.e.* the multiplication (like doubling or more) of the number of chromosomes of a species (*e.g.* by colchicine treatment of cells or tissues *in vitro*), or by allopolyploidy, involving the fusion of cells or protoplasts from two or more different species (of the same or a different genus), resulting in organisms carrying all or part of the chromosome sets of both parents (Sattler *et al.*, 2016). The latter includes techniques such as bridge crosses, embryo rescue, somatic hybridisation, translocation breeding, and others. These are traditional breeding techniques (SAM, 2017), but ploidy changes are also common in nature (*e.g.* in apple: Dermen, 1951).

Mutagenesis has been applied to induce random genetic changes by application of physical (*e.g.* irradiation) or chemical mutagens to cells, tissues or whole organisms, followed by high-throughput selection of interesting phenotypes within the resulting organisms for use for further breeding (Pacher and Puchta, 2017). Atomic gardening, as mentioned before, was one of the mutagenesis applications that was popular in the middle of the 20th century.

Genetic transformation techniques involve genetic material obtained from outside the host organism and transferred to the host using various delivery strategies, including *Agrobacterium*-mediated transformation, biolistic transformation using a so-called gene gun, microinjection (called transfection and used for animal cells), etc. The randomly incorporated sequences could be derived from a different species (in this case scientists call the outcome transgenesis), or could contain a re-arranged copy from sequences already present in the species (intragenesis) or an exact copy of sequences already present in the species (cisgenesis).

4.2 New genomic techniques and their classification

The literature search as described above did not reveal any publication of a novel 'sexual hybridisation' technique (*i.e.* involving the hybridisation of parent plants), and therefore none of such techniques is listed in the category NGTs.

Advances in science and breakthroughs in high-throughput DNA sequencing have resulted in the development of a diverse set of new **techniques for generating genetic alterations** in diverse organisms. Some of these techniques can create subtle changes in the genome in the form of single nucleotide variations (substitution, insertion or deletion of single nucleotides), while others can generate larger deletions or replacement of homologous sequences[5], depending on the technique and its way of application. Some of these new genomic techniques may also be used to insert cis-, intra- or trans-genes into the genome, but at defined locations in contrast to the genetic transformation techniques previously described, which result in random insertions.

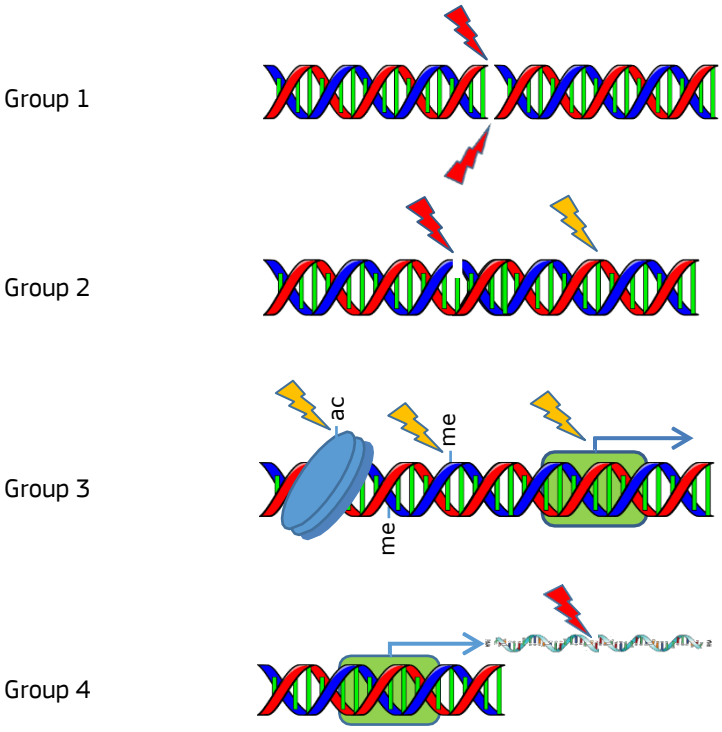
[5] Homologous sequences, in biology, are sequences that are derived from a common ancestor, and which may or may not share sequence similarity. Here the term is used in a more narrow sense to refer to sequences that share sequence similarity, in line with the concept of homologous recombination between such sequences (explained further in the text).

A defined characteristic of these NGTs is their target-specificity. That means they create changes at a specific, chosen target sequence. This makes their outcome more predictable, although unintended modifications elsewhere in the genome (so-called 'off-target effects'[6]) remain possible. The manipulation of the genome by more precise and specific changes adds a great potential to the breeding of plants, animals and microorganisms and opens up new opportunities for medical applications. A genomic alteration created by a NGT may be similar to a given natural variation occurring in the species, making it difficult if not impossible to specifically identify such alterations as man-made (ENGL, 2019).

In this study the NGTs are classified into 4 groups, differently from other classifications proposed previously (*e.g.* EFSA, 2012b; SAM, 2017). The classification used here is primarily based on the way the active components required for these techniques act on the genome, as depicted in Figure 1.

Figure 1. The four groups of NGTs classified in this report and their action on the genome

Group 1 NGTs generate a double-strand DNA break. Group 2 NGTs induce a single-strand DNA break or no DNA break. Group 3 NGTs act on the epigenetic modifications (me = methyl groups; ac = acetyl groups) of the DNA or histone proteins (blue) and/or on the transcriptional complex (green box). Group 4 NGTs act on the transcribed RNA. Interactions causing a break of the DNA or RNA are shown with red arrows, others with a yellow arrow.



The first group of NGTs (**Group 1**) is composed of diverse techniques that create a double-strand break (DSB) in the DNA. This group includes the so-called 'site-directed nuclease' (SDN) techniques, which rely on a similar set of DNA repair pathways in the cell for restoring the DSB. Other techniques in this group involve recombinases or transposases allowing the site-specific integration of a donor sequence by a concerted break and integration mechanism.

A second group of techniques (**Group 2**) is characterised by the fact that they involve a single-strand DNA (ssDNA) break or no break at all in the genome, instead of a DSB. In some cases, the induction of a ssDNA break increases the efficiency of the technique to make alterations, but this is not compulsory. Some of the NGTs in this group are derived from techniques of Group 1, but add other functionalities to them.

Techniques belonging to Groups 1 and 2 typically alter the genetic material (in the form of DNA) of the organism. Further technological advances expanded the use of NGTs to make site-specific changes in the

[6] NGTs introduce alterations at priorly defined target sequences in the genome; any changes at other locations in the genome are called off-target alterations. Some of these could be predicted through bioinformatics analysis of the whole genome sequence of the organism based on their similarity to the target site except for one or more mismatches.

epigenome[7] and on the RNA transcribed from the DNA. Such techniques often evolved from NGTs in Groups 1 and 2 and are therefore listed and described herein as well.

The third group (**Group 3**) encompasses NGTs that affect the epigenome rather than the genome. These NGTs induce alterations affecting the way the DNA sequence is read and transcribed into RNA. Also in this case, the NGTs rely on basic mechanisms of Group 1 NGTs, but add a different functionality to them.

Finally, the fourth group (**Group 4**) includes NGTs that directly act on RNA instead of DNA, creating a distinct set of NGTs. Some of these RNA-acting NGTs are also derived from one of the NGTs in Group 1.

It is important to emphasise that the NGTs described in this study are the major techniques developed since 2001 for altering the genome of plants, animals and/or microorganisms. However, the authors are aware that the list is not exhaustive and additional techniques may have been developed for use in specific organisms or for specific applications. There are overlaps between the individual techniques and variants have been developed that may fall in different groups. Advanced or alternative forms of some of the newer techniques, in particular those based on the flexible CRISPR-Cas technology, are being developed continuously and may be considered variants of existing NGTs or may receive a distinct name in the near future.

Although the possibilities to alter the genome of any organism seem endless, it should not be forgotten that the application of these NGTs generally requires a considerable amount of experimental work and optimisation for their application in a specific target organism. No NGT is universally successful with respect to every organism, and many of the reviewed studies referred to experimental work on individual cells, rather than whole organisms. The developmental processes from the individual cell of which the genome has been successfully altered through a NGT to a functioning living organism harbouring the same alteration could be long and difficult.

4.3 Nucleic acid alterations induced by the new genomic techniques

The NGTs developed in the past 20 years allow to introduce different types of alterations to the genome, epigenome or on RNA. Table 1 provides an overview of the possible alterations and the main NGTs that may be involved in generating them in the different kingdoms of living organisms (bacteria, fungi, plants and animals). The table does not specify all the functionalities of each NGT as many of them may be used in different versions and under different conditions, which cannot be all displayed in a single table. Furthermore, some NGTs may be used in combination with other NGTs for an improved performance, *e.g.* the generation of a double-strand break in the DNA by a site-directed nuclease increases the efficiency of oligonucleotide-directed mutagenesis.

The table shows that the same alterations could be generated through the application of different NGTs. Each NGT, however, has its strengths and weaknesses (as described in more detail in the next chapter). It is not the purpose of this study to compare the different NGTs and review the best approach to achieve a desired modification. This is a rapidly evolving field and the reader is referred to many excellent review papers citedin this study.

[7] The epigenome is the term used to indicate the overall chromatin composition, which marks the genome of any given cell by adding covalently linked chemical compounds on DNA or on histone proteins, thereby potentially affecting the transcription of the DNA into RNA.

Table 1. The types of nucleic acid alterations that may be induced by distinct NGTs in different organisms

Purpose	Intended sequence alteration	Type of NGT	Type of organism[1]	Donor template	Modification	Comments
Sequence correction	Substitution of one or a few bases	Base editing	P, A, F, B	No	Mostly C↔T or A↔G, with some exceptions	Additional base substitutions are possible with specific techniques. Requires a PAM, and usually changes all identical bases (*e.g.* all C's) in the targeted region
		Site-directed nuclease (SDN)	P, A, F, B	No	All base substitutions possible	Substitution is result of error-prone repair processes, which may generate random sequence variations at the targeted site, including base substitution(s)
		Oligo-nucleotide-directed mutagenesis (+ SDN)	P, A, F, B	Oligo-nucleotide (DNA or DNA/RNA)	One or a few base substitutions, defined by donor template	Oligonucleotide donor may contain one or up to 4 centrally located base mismatches, which maybe converted to the target sequence with low efficiency. Creation of a nearby double-strand break by a SDN may increase the substitution efficiency
		Prime editing	P, A, F	No DNA template, but extended guide RNA functions as RNA template	One or a few base substitutions, defined by RNA donor template	Extended guide RNA is reverse transcribed into oligonucleotide DNA template for insertion. Specific substitution of all bases possible, even when several identical bases are present in the targeted region; relaxed PAM requirement
		Site-specific recombination	P, A, F, B	Oligo-nucleotide, flanked by recombinase recognition sequences	Replacement of short target sequence	Precise replacement of target sequence, flanked by recombinase recognition sequences, with donor sequence, often leaving one copy of the recognition sequences behind
	Substitution of contiguous sequence	Site-specific recombination	B	ssDNA or dsDNA flanked by homology arms	Replacement of donor sequence	Recombineering in bacteria using recombinase systems involving several proteins (*e.g.* λ Red system)
		Site-specific recombination	P, A, F	dsDNA, flanked by recombinase recognition sequences	Replacement of donor sequence	Sequence replacement by homologous recombination at target site defined by short recognition sequences (present in the genome or inserted previously), often leading to their duplication
		Site-directed nuclease (SDN)	P, A, F, B	dsDNA template with homology arms	Replacement of donor sequence	Sequence replacement by homologous recombination, following a double-strand break at the target site
Sequence disruption[2]	Insertion of disruptive sequence in coding sequence or promoter region	Site-specific transposition	B, (A)	dsDNA transposon with transposase sequence and/or other sequences	Insertion of donor sequence	Donor template is flanked by transposase recognition sites and contains the sequence to be inserted. The transposase sequence isprovided on the transposon or on a separate plasmid. Site-specific insertion is guided by a DNA binding protein (*e.g.* Gal4, SDN) and co-introduces the transposon recognition sites
		Site-directed nuclease (SDN)	P, A, F, B	No	Deletion or insertion of random basepairs	Double-strand breaks generated by a SDN may be repaired by error-prone processes, resulting in short random deletions or insertions of 1 or a few bp, occasionally up to over 100 bp
		Site-directed nuclease (SDN)	P, A, F, B	dsDNA template with homology arms	Sequence replacement by donor sequence	Sequence replacement by homologous recombination leading to insertion of reporter or selection gene, following a single double-strand break at the target site
		Site-specific recombination	P, A, F, B	dsDNA, flanked by recombinase recognition sequences	Sequence replacement by donor sequence	Precise replacement of target sequence, flanked by recombinase recognition sequences, with donor sequence (*e.g.* reporter or selection gene), often leaving one copy of the recognition sequences behind

[1] Plants (P), Animals, including humans (A), Fungi and yeast (F), Bacteria (B)

[2] Sequence disruption may also result from a base substitution (see sequence correction) that leads to the generation of a stop codon in the protein coding sequence, or results in exon skipping

Table 1 (cont.). The types of nucleic acid alterations induced by NGTs

Purpose	Intended sequence alteration	NGT	Type of organism	Donor template	Modification	Comments
Sequence deletion	Deletion of a partial or whole gene sequence	Two site-directed nucleases or two sgRNAs	P, A, F, B	No	Sequence deletion	Use of a pair of SDNs or two sgRNAs targetted to the ends of the sequence to be deleted removes the intervening sequence
		Prime editing	P, A, F	Extended guide RNA includes a donor RNA template	Sequence deletion, defined by RNA donor template	Specific deletion of sequence up to 80 bp has been shown in human cells, relaxed PAM requirement
		Site-specific recombination	P, A, F, B	dsDNA, flanked by recombinase recognition sequences	Sequence deletion, defined by donor template	Precise replacement of target sequence, flanked by recombinase recognition sequences, with (shorter or empty) donor sequence, often leaving one copy of the recognition sequences behind.
Sequence insertion	Insertion of new sequence	Site-directed nuclease (SDN)	P, A, F, B	dsDNA template with homology arms	Sequence insertion, defined by donor template	Sequence insertion by homologous recombination (or other repair pathways), following a single double-strand break at the target site.
		Site-specific recombination	P, A, F, B	dsDNA, flanked by recombinase recognition sequences	Sequence insertion, defined by donor template	Precise insertion of donor sequence at target sequence, flanked by recombinase recognition sequences, often leaving one copy of the recognition sequences behind
		Prime editing	P, A, F	Extended guide RNA functions as RNA template	Sequence insertion, defined by RNA donor template	Insertion length limited by pegRNA sequence length (shown up to 44 bp insertion); relaxed PAM requirement
Gene regulation	Activation of gene transcription	DNA demethylation	P, A, B	No	Removal of methyl groups from gene promoter region	Demethylation may not be specific and can affect other genes
		Histone H3K27 acetylation	P, A, B	No	Enrichment of acetylated H3K27 at target site	Variable effect in different cells and different targets
		Transcription activation (CRISPRa)	P, A, F	No	H3K27 acetylation, H3K4 trimethylation	Duration of the effect is variable and not well understood
	Repression of gene transcription	DNA methylation	P, A, B	No	*De novo* addition of methyl groups to gene promoter region	Effect may not be specific and can affect other genes
		Histone H3K27 deacetylation	P, A, B	No	Removal of acetyl groups from H3K27	Variable effect in different cells and different targets
		Histone H3K4 demethylation	P, A, B	No	Removal of methyl groups from H3K4	Variable effect in different cells and different targets
		CRISPR interference (CRISPRi)	P, A, F	No	H3K9 & H3K27 tri-methylation, …	Duration of the effect is variable and not well understood
RNA sequence correction	Substitution of a single base in mRNA	RNA base editing	A	Chemically stabilised DNA oligonucleotide or sgRNA	Deamination of adenosine or cytosine in RNA	Efficiency variable and may induce significant off-target alterations. Requires continuous administration for lasting effect
	Modulation of pre-mRNA splicing	RNA splice isoform manipulation	A	No	Exon in- or exclusion from mature mRNA	Exon inclusion or exclusion depends on the positioning of the splicing effector
RNA knockout	RNA degradation	RNA interference	P, A, B	No	Cleavage of RNA	Based on Type II (Cas9), Type III or Type VI (Cas13) CRISPR-Cas systems, some of which require a PAMmer oligonucleotide. Also chemically-stabilised oligonucleotides or small interfering RNA (siRNA) may have the same effects

At the level of the DNA sequence, alterations induced by NGTs may appear as single nucleotide variations, short insertions or deletions and larger insertions or deletions. Some NGTs may have a very narrow application, *e.g.* linked to a specific base, while others could affect several types of changes depending on how they are used. Whereas the alterations are, in principle, targeted at a specific region in the genome, they can consist of short substitutions, deletions or insertions randomly created by error-prone cellular repair mechanisms, or the changes could be intentionally copied from a donor template. This donor template may contain sequences present in another individual of the same species, *e.g.* a healthy individual not harbouring a genetic disorder or a wildtype plant that is a progenitor of the cultivated elite variety. Or the donor may contain heterologous sequences from other species for their site-specific integration into the genome of a target organism, resulting in a product similar to what can be obtained by genetic transformation through EGTs. At the level of the RNA sequence, alterations that can be induced are single nucleotide variations, modulation of the splicing of pre-mRNAs and RNA degradation, including of non-coding RNAs with regulatory effects on other genes.

The most frequent DNA alterations generated by the different NGTs are summarised in Table 2. The categories used to differentiate the size ranges of the DNA alterations are chosen for maximal discrimination between the different NGTs. For instance, base editing can only change a single nucleotide in the targeted region (unless several identical nucleotides are present in the targeted sequence). Oligonucleotide-directed mutagenesis has been shown to generate substitutions, deletions or insertions up to a maximum of 4 bp, delineated by the size of the oligonucleotide used (Papaioannou *et al.*, 2012; Dekker *et al.*, 2016). Similarly, site-directed nucleases, used without donor template, generate small deletions or occasionally small insertions (usually at most 100 bp). As a result of technical limitations, prime editing could also result in deletions or insertions up to 100 bp, although alterations of this size have not been experimentally demonstrated for this rather recent technique (Anzalone *et al.*, 2019). Therefore, the size categories used in Table 3 have been solely chosen in order to discriminate the NGTs by the extent of sequence alterations they may generate, based on experimental evidence. Further justification is provided in the detailed descriptions of each NGT. As emphasised before, these size categories are dependent on the experimental conditions and target organism and are general indications that should not be taken as fixed characteristics of each of the NGTs.

Table 2. Indicative DNA sequence alterations that can be generated by NGTs, based on experimental evidence

Sequence alteration	Size of alteration	Without donor template		With donor template				
		SDN	BE	SDN	ODM	PE	SSR	SST
Substitution	1 bp	R	I	I	I	I		
	2-4 bp	(R)		I	I	I		
	> 4 bp			I				
Deletion	1-4 bp	R		I	I	I	I	
	≤100 bp	(R)/I[a]				I	I	
	> 100 bp	(R)/I[a]					I	
Insertion	1-4 bp	R		I	I	I	I	
	≤100 bp	(R)		I		I	I	
	> 100 bp			I			I	I

Abbreviations: SDN, site-directed nuclease, including homing endonucleases, zinc finger nucleases, TALEN and CRISPR-Cas; BE, base editing; ODM, oligonucleotide-directed mutagenesis; PE, prime editing; SSR, site-specific recombination; SST, site-specific transposition; R, I: random or intended sequence alteration at the target site (within brackets if possible, but rarely observed).

[a] Large deletions can result from random repair of a DSB induced by a single SDN (although less common), or can be precisely orchestrated by using a pair of SDNs targeting sites flanking the region to be deleted.

Table 3 provides an overview of the type of organisms to which each of the NGTs has been or could be applied. Many of the techniques can be applied to diverse organisms, however, this may require the use of specific versions of the general technique and the efficiency may not always be the same. Further information can be found in the chapter describing the NGTs in detail.

Table 3. Overview of the applicability of the NGTs to different types of organisms based on reported evidence

New Genomic Technique	Plants	Animals	Fungi, Yeast	Bacteria
Homing endonuclease (HE)	X	X	X	X
Zinc Finger Nuclease (ZFN)	X	X	X	
Transcription activator-like effector nuclease (TALEN)	X	X		
CRISPR-Cas[a]	X	X	X	X
Oligo-directed mutagenesis	X	X	X	X
Base editing	X	X	X	X
Prime editing	X	X	X	
Site-specific recombination	X	X	X	X
Site-specific transposition		X		X
Epigenetic modifiers	X	X		X
Epigenetic activators/repressors	X	X	X	
RNA base editing		X		
Oligonucleotide-mediated RNA interference		X		
CRISPR-Cas-mediated PAM-dependent RNA interference		X		
CRISPR-Cas-mediated PAM-independent RNA interference	X	X		X
RNA splicing alteration		X		

[a] Clustered regularly interspaced short palindromic repeats-associated protein

4.4 Understanding the impact of genomic alterations in the context of genome variability

Different organisms significantly differ in their genome sizes. For example, the genome of the bacterium *E. coli* is ~4.6 million bases, that of a fruit fly is ~180 million bases, that of a human is ~3.2 Gb (3,200 million bases). Crop species' genomes vary in size from 0.43 Gb for rice, 1.15 Gb for soybean, 2.4 Gb for maize, to 17 Gb for wheat.

The representation of the genome as a species-specific reference sequence of DNA has radically changed as a result of whole genome sequencing (WGS) projects since 2003, when the first human reference genome was published. The feasibility of WGS analysis is supported by next generation sequencing (NGS) technologies, which require substantial computational and biostatistical resources to acquire and analyse large and complex sequence data. The NGS techniques have further evolved over the past 10 years, both in terms of accuracy, read-length, speed and costs. This development created enhanced opportunities for the sequencing of polyploid species, like many agri-food crops, in which the assembly of the data into chromosomes has been challenging (Kyriakidou *et al.*, 2018).

Reference genomes are often incomplete and do not represent the full range of genetic diversity. The sequencing of many individuals of a species, such as in the 1000 genomes reference project (Auton *et al.*, 2015), has confirmed that genome sequences could differ considerably between any two individuals of a species. For instance, the typical difference between the genomes of two human individuals was estimated at 20 million base pairs or 0.6% of the total of 3.2 billion base pairs. More than 99.9 % of the differences are single nucleotide variations and short indels (insertions and deletions < 50 bp), but, in addition, a typical human has 2,100 to 2,500 structural variations, which include approximately 1,000 large deletions, 160 copy-number variants, 915 Alu (transposable element) sequence insertions, 128 L1 (transmembrane protein family) sequence insertions, 51 SVA (retrotransposon) insertions, 4 NUMTs (nuclear transpositions of mitochondrial DNA), and 10 inversions. On a species level, 84.7 million single nucleotide polymorphisms (SNPs), 3.6 million short insertions/deletions (indels), and 60,000 structural variants have been characterised during sequence analysis of 2,504 human genomes (Auton *et al.*, 2015). Similarly, WGS of other animal species has shown the presence of variants resulting from natural and breeding selection. For instance, in pigs, *de novo* sequencing of representatives from different regions uncovered a substantial number of novel SNPs (>6 million out of over 33 million SNPs identified in total) and structural variants, as well as 137 Mb sequences (divided over close to 84,000 sequences of ≥500 bp) harbouring 1,737 protein-coding genes that were absent in the reference genome assembly (Li *et al.*, 2017).

Larger genome size, polyploidy, presence of highly repetitive genomic regions, and genome duplication are the major challenges for sequence analysis of most of the plant genomes. For example, the repetitive fraction of the human genome varies between 35 % and 45 % of the genome, whereas in maize it is 64–73 % (Imelfort and Edwards, 2009) and in wheat even 85 % (International Wheat Genome Sequencing Consortium, 2018). Over 300 vascular plant species have been sequenced up to now, but only a fraction of these have been fully assembled and annotated (Kersey, 2019). By 2023, at least one reference genome is expected to be available for 10,000 plant species (Cheng *et al.*, 2018). A reference genome, however, does not represent the enormous genome variability of a species. Dispensable genes, absent from a portion of plants within the species, may constitute a significant portion of the overall pan-genome, *e.g.* around 20 % in soybean (Li *et al.*, 2014). A comparison between two maize inbred lines showed that their genomes contained 3,408 and 3,298 unique indels, respectively, with an average size of approximately 20 kb (20,000 base pairs) and a range covering 1 kb to over 1 Mbp (Jiao *et al.*, 2017). Another example is hexaploid bread wheat, composed of A, B and D subgenomes, for which the first reference genome took 13 years to complete. To provide a flavour of its complexity: almost 4 million copies of transposable elements belonging to 505 families were annotated, >100,000 high-confidence (HC) protein-coding loci were identified and >160,000 other protein-coding loci were classified as low-confidence (LC) genes, representing partially supported gene models, gene fragments (pseudogenes), and orphans. Of the bread wheat HC genes, 27 % (almost 30,000 genes) are present as tandem duplicates (International Wheat Genome Sequencing Consortium, 2018). In this context, the Ensembl platform (http://www.ensembl.org/), together with the EU-funded transPLANT infrastructure (http://www.transplantdb.eu/), collect and compile sets of variations for different crop plants, such as barley, tomato, wheat. Another example is the 3,000 Rice Genomes Project (2014), which includes rice data derived from a global, public genomic database of 3,024 different rice cultivars.

Spontaneous natural mutations change the genome at each reproduction cycle. For instance, in the model plant *Arabidopsis* (*Arabidopsis thaliana* (L.) Heynh.), there is a seven in 1 billion chance that any given base pair will mutate in a generation (Ossowski *et al.*, 2010), meaning that 175 new variants (SNVs) would arise per 100 individual plants per generation. In rice, more than 54,000 novel DNA sequence variants were identified in a line that went through *in vitro* culture (and 8 cycles of self-fertilisation), compared to the wild-type line, without showing any different phenotype under normal growing conditions (Zhang *et al.*, 2014a).

The relatively slow rate of natural mutation is increased by several orders of magnitude by induced mutagenesis, using irradiation or chemical treatment of seeds or pollen, which has been applied to plant breeding since the mid-nineties (Anderson *et al.*, 2016). Such mutant plants have been incorporated in traditional breeding programmes and have contributed to the current crop diversity. The DNA alterations induced by conventional mutagenesis can be extensive. For instance, 41 mutant plants resulting from fast-neutron (FN) irradiation in rice were sequenced and the sequences compared to the parental line: a total of 2,418 FN-induced homozygous and heterozygous mutations were detected in these 41 rice lines, including 1,273 single base substitutions, 864 deletions, 145 insertions, 82 inversions, 49 translocations, and five tandem duplications. The sizes of FN-induced insertions and deletions range from 1 bp to 678 kb, with small (≤10 bp) deletions and insertions accounting for 73.2 % of all deletion and insertion events. Per individual line, on average 59 mutations were detected, affecting on average 31 genes. The number of mutations induced by FN irradiation was about 6 times more than the spontaneous mutation rate (10 mutations per generation) in rice (Li *et al.*, 2016a). In another study, approximately 390 single base substitutions per line were detected in an EMS (ethyl methanesulfonate) mutant rice collection (Henry *et al.*, 2014).

NGTs generate alterations in the context of the highly diverse and evolving genome sequence of a species. Whereas some of the NGTs may be used to insert genetic material derived from another species into the genome, the outcome of many of the NGTs is similar to natural variations within the endogenous genetic material of an organism. These variations can be small, *e.g.* a single basepair substitution, deletion or insertion, or large, including deletions or insertions of over 1 kb. On a genomic level, these alterations are usually small compared to the genetic variability of genomes described above.

Some of these modifications may reconstruct a sequence that naturally exists in another individual, strain, variety, subtype or culture of a species. For instance, the main cause of childhood blindness (Leber congenital amaurosis) is a mutation in a gene that results in a splicing defect. This genetic disorder may be cured (*i.e.* changed back to the normal sequence) in humans carrying the genetic disorder through genome editing (See Nature 579, 185 (2020), doi: 10.1038/d41586-020-00655-8). Similarly, genes for higher yield or product quality that have been lost or became inactive through breeding may be re-introduced into elite germplasm at the original location using genome editing. Other modifications may delete or inactivate undesirable genes, thereby producing crops with enhanced or modified food quality, such as low acrylamide potatoes, allergen-free peanuts or gluten-free wheat (Sedeek *et al.*, 2019). In fact, much of our current knowledge on the function of genes and their relationship with a phenotype can be harnessed to generate genetic alterations and to create new phenotypes across species. For instance, spontaneous knockout mutations in the *mlo* gene provided broad-spectrum resistance against a bacterial disease (powdery mildew) in barley, tomato and *Arabidopsis*. However, similar mutations have not been identified in wheat *mlo* genes. Genome editing with TALEN resulted in the simultaneous knockout of all three *mlo* homologs in bread wheat as a result of mostly 1-10 bp deletions. This conferred durable resistance against this devastating pathogen (Wang *et al.*, 2014).

Another approach for looking at genome diversity is to evaluate the probability of whether a sequence is unique in the context of a whole genome. While genome sizes can vary by orders of magnitude, the calculated theoretical probability that a random sequence is unique in the genome of various crops boil down to a consistent relatively narrow size range between 19 and 21 bases. This was also confirmed by analysing the published (relatively small) rice genome sequence (JRC, unpublished results). It is expected that unique sequences in larger and polyploid genomes, often containing a high proportion of repetitive DNA, will be longer in size compared to those calculated above. A NGT that generates a modified sequence of a size smaller than the theoretical size for 'uniqueness' determined for the species may already occur elsewhere in the genome and would therefore not be unique. Such an argumentation is based on probabilities and may disregard the reality of the evolving genetic diversity of genomes. In some cases, a sequence larger than the theoretical threshold for uniqueness may already occur in the genome of the species and is therefore not unique, *e.g.* when a missing sequence resulting in a disease has been reverted back to the non-disease state.

Conversely, 'nullomers' are short sequences of DNA that do not occur in the genome of a species (for example, humans), even though they are theoretically possible. For example in rice (*Oryza sativa*) the 12 bp sequence TTAATGATCCGC is considered to be a nullomer (see https://www.nullomers.org/) although statistically it should be present around 30 times in each genome of *Oryza sativa*. Therefore, if a stretch of *Oryza sativa* is altered to become "TTAATGATCCGC" this could be considered with the current knowledge as not occurring in nature. However, 12 bp oligomers with a different sequence are present in the genome of this species.

4.5 Delivery of the active components for the NGT to the target cells

The NGTs described in this study (just as it was the case for EGTs) require delivery systems to bring the active components for genome alteration into the targeted cells. The delivery systems differ between plant, animal and microbial cells and some of them may be specific for a certain organism.

In **plants**, the use of *Agrobacterium tumefaciens* and its T-DNA are widely employed as vehicle for transfer and (stable) integration of the transgenes (Watanabe *et al.*, 2016). Also taxonomically related rhizobacteria could be made competent for gene transfer to plants (Broothaerts *et al.*, 2005). Other approaches are based on biolistic systems shooting metal particles with bound DNA at high velocity into the cells (also called microparticle bombardment). In some cases, direct uptake by protoplasts of recombinant DNA from culture media, supported by electroporation or polyethyleenglycol (PEG), has been successful. Several viruses have also been employed for delivery of DNA or RNA (Marton *et al.*, 2010; Ma *et al.*, 2020). More recent developments are the use of nanomaterial-based delivery systems, like carbon nanotubes or mesoporous silica nanoparticles (MSNs), which may overcome host-range limitations (Wang *et al.*, 2019a).

Once the integrated transgene DNA expressing the active components for the NGT (or EGT) have generated the altered DNA sequence or epigenomic effect, they may be segregated in the next generation through sexual crossing or through the application of a site-specific recombination system like Cre-LoxP.

Alternatively to transgene integration, the genome alteration components could be transiently expressed from a recombinant DNA plasmid or vector, following biolistic delivery or *Agrobacterium*-mediated transformation (Clasen *et al.*, 2016; Veillet *et al.*, 2019). In this case, the introduced DNA is expressed into proteins and/or RNA and subsequently lost due to endogenous nuclease activity.

Additionally, the components could be delivered as protein or ribonucleoprotein complex (protein plus RNA) or as messenger RNA (mRNA) rather than DNA, which could be realised by PEG-mediated transfer to protoplasts or microparticle bombardment (Woo *et al.*, 2015; Kim *et al.*, 2017a; Liang *et al.*, 2017a; Ran *et al.*, 2017; Malnoy *et al.*, 2016; Li *et al.*, 2019b; Alok *et al.*, 2020). The active (ZFN) protein has also been transfected to plant cells (wheat microspores) with the help of cell-penetrating peptides (Bilichak *et al.*, 2020).

Some NGTs (*e.g.* oligonucleotide-directed mutagenesis) rely on the provision of only a single oligonucleotide or double stranded DNA donor template, which are usually delivered by particle bombardment and act transiently before being inactivated by cellular nucleases (Zhu *et al.*, 1999; Rivera-Torres *et al.*, 2016).

A special case of intergeneric wide-cross delivery of genome alteration components has been developed in wheat. In this case, maize pollen that expressed the genome editing components for editing a wheat gene, was applied *in vivo* to wheat spikelets. Following culturing of the embryos *in vitro* a low number of genome edited wheat haploids were obtained (Kelliher *et al.*, 2019). The maize inducer line was used herein both as haploidy inducer and delivery vehicle for the editing components, whereby the maize chromosomes were naturally eliminated during this process.

The application of NGTs to **animals** usually involves one of various transfection methods for plasmid vectors or oligonucleotides, applied to diverse types of cells or tissues that are able to regenerate into an organism (Beurdeley *et al.*, 2013; Merkle *et al.*, 2019). Messenger RNAs may be cytoplasmically injected, *e.g.* into embryos, or electroporated (Carlson *et al.*, 2012) and ribonucleoproteins may be electroporated or introduced with the help of lipid nanostructures or gold particles (Liang *et al.*, 2015). Oligonucleotides may be injected intravenously or at different sites, depending on the target tissue and function (Tsoumpra *et al.*, 2019). Alternatively, viral delivery systems, based on the adeno-associated virus (AAV) or lentivirus, may be employed and are commonly used for plasmid delivery (Holkers *et al.*, 2013; Liu *et al.*, 2017a; Carpenter *et al.*, 2019). The size of the genome alteration components is, however, a limiting factor for viral packaging.

Genome editing in **fungi** generally employs the uptake of the NGT tools through physical methods such as electroporation or bombardment (Morio *et al.*, 2020) or is aided by chemicals like lithium acetate or polyethyleneglycol (Hassan *et al.*, 2020; Bae *et al.*, 2020). The diversity of fungal and yeast species require different delivery strategies and different NGTs to achieve editing (Song *et al.*, 2019; Morio *et al.*, 2020).

The genome of **bacteria** could be modified by known transformation methods such as direct uptake of exogenous DNA from the medium, aided by electroporation (Corts *et al.*, 2018), conjugation (Wendt *et al.*, 2016) or transduction involving a bacteriophage virus (Sharan *et al.*, 2009). DNA encoding the genome alteration components from a plasmid vector is generally only transiently present, but could also be stably inserted in the bacterial genome.

4.6 Differences between new and established genomic techniques

In this section the major differences between the most widely used EGTs, *i.e.* genetic transformation techniques, polyploidy induction and traditional mutagenesis, and the NGTs described above are reviewed.

Random versus site-specific alterations: One of the major differences between EGTs and NGTs, and a driver for research and development toward new technological approaches, is the ability for site-specific alteration of the genome. The EGTs often rely on difficult to control random changes and require significant investments in time and resources for the further testing of the resulting organisms. For instance, traditional mutagenesis experiments in plants and microorganisms induce thousands of unknown mutations and require large-scale genetic analysis, field testing over several generations and selection of those individuals that display the desired characteristics without any apparent unintended effects (Parry *et al.*, 2009; Sikora *et al.*, 2011; Irshad *et al.*, 2020). Other examples are the transformation or transfection technologies in plants and animals, where an unwanted integration of the transgenes into endogenous genes or transcriptionally silenced regions may give unexpected effects and should be avoided (Schubert *et al.*, 2004; Rajeevkumar *et al.*, 2015). The site-specific integration of exogenous DNA by homologous recombination is instead common in yeast and bacteria, but very rare in higher organisms (Oldenburg *et al.*, 1997).

In contrast, NGTs allow, in principle, to make changes in the genome at one or more precise location. These locations can be carefully chosen as target sequence(s) in the genome of the specific cell(s) used as subject of the NGT experiment, based on the correct knowledge of the genomic sequence of that particular organism. The target location(s) can be chosen to be sufficiently unique in the genome to avoid changes at off-target

sites sharing a limited extent of sequence similarity with the target site. Although the occurrence of unwanted off-target alterations following the use of NGTs is often not negligible and needs careful evaluation during the design of the experiments, the frequency of such modifications is generally much lower compared to the range of potential unintended effects resulting from the use of EGTs, in particular compared to traditional mutagenesis (Anderson *et al.*, 2016; SAM, 2017).

NGTs could also be used in combination with a donor sequence to insert cis-, intra- or transgenes at a defined location in the genome (d'Halluin *et al.*, 2019; Zhu *et al.*, 2017; Bi *et al.*, 2018; Collier *et al.*, 2018).

Unexpected versus predictable outcomes: In traditional transformation and induced mutagenesis, the outcome is largely unpredictable because the number and position of the mutations cannot be controlled. The effect of interspecies or intergenus crossing leading to doubling of the chromosomes is also unpredictable due to the interactions between the new genes and chromosomal translocations or deletions which may consequently result (Zhang *et al.*, 2013a). Also in transformation experiments unintended effects could occur due to the insertion of the transgenes into endogenous gene sequences (coding or regulatory), affecting their expression, or the insertion of several copies of the transgenes leading to unwanted gene silencing (Schubert *et al.*, 2004; Rajeevkumar *et al.*, 2015), or chromosomal translocations (Jupe *et al.*, 2019).

NGTs target specific sequences in the genome: The targets are often sequences where the effect of the change is known from other organisms, *e.g.* wild varieties, displaying the new sequence together with the expected phenotype. The outcome of such experiments is therefore more predictable, in any case for monogenic traits. As a consequence, the time to develop a genome edited organism expressing the desired phenotype is generally much shorter than for any of the EGTs and the costs involved are lower (disregarding the regulatory requirements; SAM, 2017). Bioinformatics tools are available to prevent or limit the risks of a NGT experiment to affecting off-target sequences (Cui *et al.*, 2020). Nevertheless, similarly as many unintended changes could go unnoticed after application of EGTs, some off-target modifications may result from NGT use. Depending on the type of alteration, its location and the organism, such additional modifications may or may not have a phenotypic effect (Weber *et al.*, 2012; SAM, 2017).

The different gene pools for making genomic alterations: Transgenesis as EGT (and allopolyploidy resulting from combining the genomes from different species) may be used to express sequences that are derived from unrelated species. In contrast, traditional mutagenesis, cis/intragenesis and autopolyploidy induction are restricted to the existing endogenous gene pool to induce alterations. The same is valid for the different types of NGTs, except when used in combination with a heterologous donor sequence for homologous recombination (HR). This means that no new exogenous sequences are inserted into the genome of an organism via NGTs and that all sequence alterations are derived from endogenous sequences or from random insertions or deletions generated by the endogenous repair systems.

HR with a homologous donor sequence for the site-specific gene correction or replacement also plays within the endogenous gene pool. Notwithstanding this, a donor template intended for gene correction, *e.g.* for generating disease resistance, may contain a sequence that is only found in another species. This means that the gene pool may be broadened compared to what is possible through sexual hybridisation.

The situation is different when a heterologous donor template is employed for insertion of a sequence not known to occur in the species' gene pool or in the genome of a cross-compatible species. The latter can be called 'site-specific transgenesis' and extends the genome modification options to sequences from any unrelated organism.

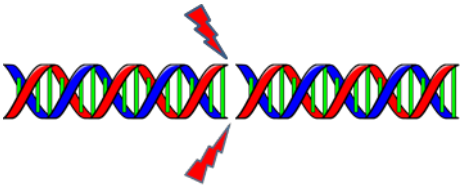
The different type of genomic alterations: As described in section 4.1, genetic transformation techniques involve genetic material obtained from outside the host organism and transferred to the host using various delivery strategies. The changes induced by these techniques largely consist of insertions of new sequences. NGTs on the other hand could be employed to make small or large deletions of sequences, short insertions or basepair replacements (gene correction). Many of the NGT-generated alterations are limited in size to the single nucleotide level or to a number of consecutive or dispersed basepair changes, much like natural genomic variations. For instance, CRISPR-Cas-mediated knockouts in two genes in tomato resembled natural white flower and yellow flesh mutants, although the edited alterations (SNPs and short indels) are much smaller than the natural mutations (d'Ambrosio *et al.*, 2018). Such directed short insertions or deletions of any size are impossible to be achieved by many of the EGTs, although natural and traditional mutagenesis may create similar changes but in a random manner (Weber *et al.*, 2012; Anderson *et al.*, 2016). For instance, EMS mutagenesis in flax resulted in SNP variants in an UDP-glucosyltransferase gene (Fofana *et al.*, 2017), a type of alteration that could also have been obtained by applying a NGT.

Like transformation techniques based on EGTs, NGTs can also be employed for sequence insertions when combined with a co-delivered donor template, which is then directed to a defined location in the genome.

Singleplex versus multiplex: Traditional mutagenesis and polyploidy induction generate multiple alterations in the genome. Several of the NGTs, *e.g.* the CRISPR-Cas based techniques, can also be employed in the form of multiplex approaches, introducing alterations at different sites in the genome simultaneously. Transformation technologies generally intend to make one alteration per experiment, and re-transformation or sexual crossing is needed to stack multiple alterations into the same genome (Halpin, 2005). The number of changes induced by NGTs is therefore generally much lower compared to traditional mutagenesis experiments, but comparable to or, in the case of multiplex editing, generally higher than in transformation experiments.

Detectability: the insertion of a new DNA sequence into the genome through genetic transformation technologies creates a novel junction sequence between the insert and the endogenous genome. These novel, unique junctions can be targeted by PCR-based detection and identification techniques, which also allow quantification of the GMO content in a food or feed mixture. A site-specific insertion of heterologous DNA by NGTs can generally be detected and quantified by the same approach and technology. In the case of alterations encompassing the endogenous gene pool the specific PCR-based detection of the edited product may be more challenging, depending on the size of the alteration (SNP versus long deletion), and whether it has created a distinctive junction in the genome that can be targeted by the PCR technique. Detection also depends on the prior knowledge of the altered sequence and on the kind of product to be tested for the presence of the genome-edited organism, *i.e.* alterations detected in individual organisms may be informative of genome editing, but mixtures of several organisms may be impossible to evaluate in the same sense. When the same sequence alteration occurs in another variety of the same species, or even in another organism, a mixed product may equally be positive due to the presence of the sequence-edited organism or of the non-edited organism that displays the same sequence alteration, and the two cannot be distinguished in (food or feed) mixtures (further details are reviewed in ENGL, 2019; Grohmann *et al.*, 2019).

5 Group 1: Genome editing involving DNA double-strand breaks (DSBs)



A DSB is simply a break in duplex DNA. Creating breaks and rejoining double stranded DNA is fundamental to all living organisms. DSBs consist of two double-strand DNA ends at a single location facing each other. DSBs are natural phenomena in cells and may be generated by endogenous factors such as DNA replication, DNA recombination, DNA transposition, by products of metabolism (*e.g.* reactive oxygen species) or by exogenous factors such as UV, ionizing radiations or chemical mutagens (Manova and Gruszka, 2015). Some of the processes that induce DNA DSBs involve endogenous endonucleases or other enzymes that cut DNA.

DSBs occurring in the genome are dangerous to the cell as they may lead to cell death if not repaired rapidly. Cells have thus evolved several repair pathways to mend the DNA. These repair pathways have been effectively co-opted for genome editing following the induction of a DSB by one of several site-directed nucleases (SDNs). This forms the principle of the widely used NGTs that will be described first. In contrast, recombinase and transposase-mediated mechanisms rely on concerted mechanisms for DNA break and rearrangement without requiring DSB repair pathways.

5.1 Site-directed nuclease-mediated genome editing

Genome editing in its most widely known concept starts with recognition of a target sequence and creation of a double strand break (DSB) in the DNA. A range of molecular tools for inducing DSBs at specific sites in the genome are available, each with their own characteristics, *i.e.* homing endonucleases, zinc finger nucleases (ZFN), transcription activation-like effector endonucleases (TALEN) and CRISPR-Cas. These nucleases act sequence-specifically and are generically called site-directed nucleases (SDN).

The DSB approach for genome editing follows a two-step process, *i.e.* recognition and binding of the SDN to a target sequence and generation of a DSB, and subsequent repair of the DSB by endogenous cellular pathways. In this section, the different SDN techniques employed for the generation of DSBs are described first (Section 5.1.1). The repair of these DSBs follow pathways that are independent of the technique used to generate them. These repair pathways are described in a separate subsection (Section 5.1.2) and are similar for all SDN techniques.

From a conceptual point of view, SDN applications have been categorised into SDN-1 (no donor sequence template provided, random repair of the DSB), SDN-2 (template provided for homology-directed repair, resulting in small, precise alterations) and SDN-3 (template provided for homology-directed insertion of large sequences like genes) (EFSA, 2012b; SAM, 2017; Pacher and Puchta, 2017). While such a grouping may be informative for regulatory purposes, it does not reflect intrinsic characteristics of the different SDN techniques and cannot be used for their description. From the review presented here it will become clear that many techniques described here may be used in the form of SDN-1, 2 or 3. Moreover, repair pathways typical for one SDN category may induce alterations representative of another one. For example gene addition (SDN-3) may also occur through the error-prone random non-homologous end joining repair pathway typical for SDN-1. Also alternative repair pathways, inducing errors or not, can interfere during the repair of nuclease- induced breaks. Furthermore, the provision of a homologous DNA template is not always necessary to induce homology-directed repair, as this template can also be generated inside the cell from a dual-function guide RNA or RNA may function as template (see Prime editing in Group 2 NGTs). As the mechanisms at the basis of the different techniques are progressively being elucidated, endless variations may be developed for specific purposes, making categorisation more complex.

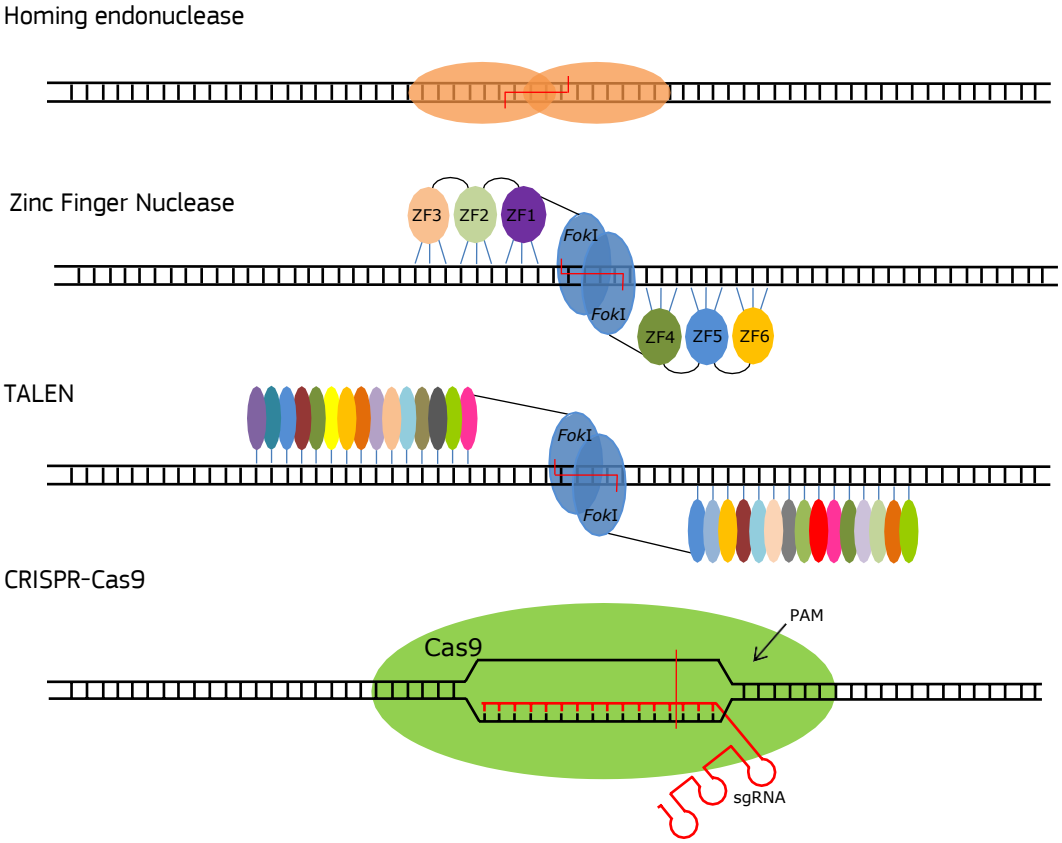
5.1.1 Site-directed nucleases generate a double-strand break in the DNA

Several classes of natural or designed nucleases have been employed as tools for targeted gene modification. The first generation of SDNs, including homing endonucleases, zinc finger nucleases (ZFN), and transcription activator-like effector nucleases (TALENs) recognise DNA exclusively through protein-DNA interactions, whereas the most recent member of the SDN group, the CRISPR/Cas9 system, also depends on a targeted

RNA-DNA interaction (Figure 2). The characteristics of the four major SDN techniques which have been applied for genome editing in various organisms are described below.

Figure 2. General mode of action of the different classes of site-directed nucleases (SDNs).

(A) Homing endonucleases are single monomers with integrated DNA binding and nuclease functions for site-specific cleavage of DNA leaving 3' overhangs (red line). (B) Zinc Finger Nucleases (ZFNs) are chimeric proteins consisting of 3 or 4 zinc finger domains, each recognising three nucleotides, fused with a nuclease (*e.g.* FokI). Two such ZFNs work in tandem for the cleavage of DNA leaving 5' overhangs. (C) Transcription activator-like effector nucleases (TALENs) are fusions of a nuclease domain (*e.g.* FokI) with a number of TALE proteins that each recognise a single nucleotide. Two TALENs dimerise to cleave DNA, leaving 5' overhangs. (D) The CRISPR-Cas9 system consists of a Cas9 nuclease forming a complex with a single-guide RNA (sgRNA), which cleaves the DNA at 3 bp from the PAM (protospacer adjacent motif), generating blunt ends (red vertical line) (modified from Caroll, 2014).



5.1.1.1 Homing endonucleases

Introduction: Homing endonucleases (HEs) or meganucleases are considered mobile introns with large DNA recognition sequences, whose function in nature is to initiate DSB-induced DNA recombination events in a process referred to as homing (Chevalier and Stoddard, 2001). Homing is the lateral transfer of an intervening sequence of a gene (either an intron or intein[8]) to a homologous allele that lacks the sequence, using homologous recombination. HEs are found in all branches of life and several hundreds of HEs have been identified in bacteria, archaea and in the nuclear, chloroplast and mitochondrial genomes of eukaryotes (Jurica and Stoddard, 1999). They were the first site-specific nucleases employed for genome editing in plants and mammalian cells. The HE technique is not widely employed because the restricted sequence-specificity of these nucleases hinders their widespread use and engineering HEs with new specificities is challenging (Smith *et al.*, 2006).

Mechanism: Genome editing using HEs requires the HEs to be expressed in the cells following the use of a transformation technology (in plants), transfection or transduction (in animal cells), or other means leading to

[8] Introns are sequences in the genome that are post-transcriptionally excised from the pre-mRNA, thereby linking the 'exons' to each other. Inteins are sequences that are transcribed and translated into a protein, but post-translationally excised from the amino acid sequence, thereby linking the 'exeins' together. The sequences for HEs are often found within intron or intein sequences.

the uptake of the DNA encoding the HE. The HE can be transiently expressed from a DNA construct or stably inserted into the genome of the host (and potentially segregated away lateron). The different HEs recognise and bind target DNA sites of between 14 and 40 bp and generate an asymmetrical DSB in the DNA through their endonuclease activity. As reviewed in more detail below (Section 1.2), these DSBs may be repaired by the so-called 'non-homologous end joining' (NHEJ) repair pathway, re-ligating the broken ends seamless or generating small nucleotide alterations (indels[9]). In the presence of an exogenous template sequence, repair may occur through homologous recombination (HR), similar to the function of natural HEs. The presence of DNA overhangs following a DSB generated by a HE increases the frequency of HR of the target sequence. Conversely, the addition of DNA end-processing enzymes that trim the overhangs together with a HE has shown to increase NHEJ editing frequency in human cells (Delacôte *et al.*, 2013).

Homing endonucleases (HE): detailed mode of action

On the basis of sequence and structural motifs, HEs are divided into five families: GIY-YIG, HNH, His-Cys box, PD-(D/E) XK, and the best-known group of the LAGLIDADG HEs (Zhao *et al.*, 2007). Two properties distinguish HEs from other site-specific endonucleases like restriction endonucleases. Most importantly, HEs recognize and bind exceptionally long target sites (14–40 bp), despite their relatively small size (<40 kDa). Furthermore, HEs tolerate subtle changes in target site sequence. Many HEs (*e.g.* I-*Cre*I, recognising 22 bp) act as homodimers, although others act as monomer (*e.g.* I-*Sce*I, recognising 18 bp). The DNA binding and cleavage functions of LAGLIDADG HEs are intertwined in a single protein domain with an αββαββα fold (Chevalier and Stoddard, 2001). The two β-strands create a saddle-shaped DNA binding region mimicking the helical twist of the DNA (Silva *et al.*, 2011). Some HEs have been engineered for a wider range of binding and cutting sites or have been used in combinations to recognise multiple sites (*e.g.* Gao *et al.*, 2010; d'Halluin *et al.*, 2013).

Following DNA binding an asymmetrical DSB cleavage occurs through the endonuclease activity of the HE, leaving 3'-overhangs of typically 4 nucleotides.

Inducible modifications: Homing endonucleases induce a DSB, which is predominantly repaired through NHEJ, often generating small insertions or deletions of a few nucleotides. In maize, an engineered I-*Cre*I has been used for targeted modification of the *liguleless1* gene promoter, resulting mostly in deletions (2-220 bp) at the target site, and a few plants also revealed short insertions (Gao *et al.*, 2010). The co-delivery of a donor template with long homology arms may result in sequence exchange between target DNA and donor sequence through HR. For instance, homology arms of 1,561 bp upstream and 2,072 bp downstream of the target site were successfully used to insert two transgenes in cotton (d'Halluin *et al.*, 2013).

Result: Without co-delivery of a donor template, the use of a homing endonuclease may result in a gene knockout through the generated codon frameshift resulting from the random indels created at the target site. When a donor template with homology arms is co-delivered with the HE, correction of a malfunctioning gene or replacement of an endogenous gene by a desired coding sequence are possible. In addition, insertion of a new sequence (cis/intragenesis or transgenesis) may be achieved, *e.g.* for the stacking of genes at a safe harbour[10] in the genome (d'Halluin *et al.*, 2013).

Target organisms: The HE technique has mainly been used in research for the development of mammalian cell lines (Cabaniols and Pâques, 2008), or for studying the process of homologous recombination in yeast, bacteria, mice and plants (Gouble *et al.*, 2006; Watanabe *et al.*, 2016).

Known off-target effects: As for any technique generating a DSB off-target modifications are possible depending on the conditions employed. Due to the limited use of HEs for genome modification there is not much information on the occurrence of off-target alterations. In general, due to their long DNA recognition sequences, the probability for inducing alterations outside the target region is relatively low.

Limitations and gaps in knowledge: The long recognition site of natural HEs restricts their utility for genome editing because of the scarcity of targetable sites in the genomes of host cells. To compensate for this, researchers have engineered these nucleases for a wider range of binding and cutting sites or they have used combinations of different homing endonucleases to recognise multiple sites (Chevalier and Stoddard, 2001). Using a mutagenesis and high-throughput screening approach separate DNA-binding subdomains have been generated in I-*Cre*I (heterodimeric) variants with locally altered substrate specificity (Smith *et al.*, 2006), as

[9] The term 'indels' is commonly used to refer to insertions and/or deletions in the form of acquisitions or loss of basepairs from a target sequence, often occurring in a random manner during repair of a DSB.

[10] Safe harbours or hot spots in the genome are regions that have been experimentally shown to allow transgene insertion and resulting in expected transgene expression without disrupting endogenous genes or creating (off-target) effects on the expression of nearby genes.

well as in several other HEs. However, the engineering of HEs for wider applications is currently still considered as inefficient, laborious and time consuming.

5.1.1.2 Zinc Finger Nucleases

Introduction: Zinc Finger Nucleases (ZFNs) are hybrid proteins formed by linking the DNA cleavage function of an endonuclease to a class of DNA binding proteins called zinc fingers. These zinc fingers were initially identified in diverse genes of *Xenopus*, Drosophila and yeast based on their characteristic loop-like domains, the 'fingers', and their metal-binding characteristics (Vincent, 1986). The zinc finger domains have subsequently been identified in unrelated genes from many eukaryotic organisms and share a common DNA binding motif conserved throughout evolution. The modular structure of the zinc finger domains has been exploited as a genome editing tool by linking different zinc fingers, each recognising a desired DNA triplet target sequence, to the cleavage domain of bacterial endonucleases like *Fok*I (Kim *et al.*, 1996). Such ZFNs may be engineered to potentially recognise and cleave any target sequence in the genome and have been used for genome editing in diverse organisms.

Mechanism: ZFN are engineered proteins consisting of a DNA binding domain formed by in tandem arrayed zinc fingers linked to the catalytic domain of a DNA endonuclease, usually *Fok*I. Each zinc finger interacts with a triplet code (3 subsequent bases) of the DNA sequence and different zinc fingers may be combined to target a specific DNA sequence (Figure 2). ZFNs usually contain 3 or 4 fingers. Since a three-finger ZFN requires two copies of the 9 bp recognition sites in a tail-to-tail orientation in order to dimerise and produce a DSB, it effectively has an 18 bp recognition site, which is long enough to specify a unique genomic target site in plants and mammals (Durai *et al.*, 2005). In contrast to their long recognition sites, each finger in reality often recognises only two bases out of the cognate DNA triplet, reducing the presumed specificity. In some cases, four-finger ZFNs have been shown to be both more specific and less toxic to the cells. Toxicity of ZFNs has been reported repeatedly in the literature and seems to result from ectopic cleavage of genomic sites. Also the use of obligatory heterodimerising ZFNs has reduced the cytotoxicity of ZFNs (Miller *et al.*, 2007). The delivery of the ZFNs as protein instead of a DNA construct that is expressed inside the cells has also proven advantageous in mammalian cells (Gaj *et al.*, 2012; Bilichak *et al.*, 2020).

Upon recognition of the target site and binding of both ZFN units of the dimer to the DNA, an assymetrical DSB is generated, generating 5' overhangs of 4 or 5 bp (Smith *et al.*, 2000). This DSB is repaired by any of the DNA repair mechanisms described below. When an exogenous repair template is also supplied then sequence alterations encoded in this donor may be incorporated into the genome by homology-directed repair.

A useful alternative to ZFN is the construction of ZF nickases by inactivation of the catalytic activity of one monomer of the ZFN dimer (Ramirez *et al.*, 2012). Such nickases generate single-strand DNA breaks which pushes the balance in favour of homologous recombination instead of NHEJ in plants (see Section 5.1.2). Alternative fusion proteins with different functionalities besides nuclease-mediated cleavage of DNA have also been constructed between zinc fingers and recombinases, methylases, and transcriptional activators and repressors (Durai *et al.*, 2005; reviewed in various chapters below).

Zinc finger nucleases (ZFN): detailed mode of action

The zinc finger repeats are typically 30 amino acid long domains folding into a ßßα structure with a consensus sequence (Tyr,Phe)-X-Cys-X^{2-4}-Cys-X^{3}-Phe-X^{5}-Leu-X^{2}-His-X^{3-5}-His, with X being any amino acid and the numbers (or ranges) referring to the number of X amino acids (Berg, 1988; Kim *et al.*, 1996). Each of these sequences binds a zinc (II) ion to form the structural domain termed a zinc finger. These proteins, like many sequence-specific DNA-binding proteins, bind to the DNA by inserting the α-helix into the major groove of the DNA double helix. The key feature of the DNA-binding zinc fingers is that they are composed of independent units linked by flexible peptidic joints: the structurally conserved pairs of histidine and cysteine residues bind to a zinc ion for structure stabilisation, while the phenylalanine and leucine residues form a hydrophobic core within each finger. The fingers bind to nucleic acids by interaction of the positively charged DNA-binding amino acids with the phosphate backbone of DNA (Vincent, 1986; Berg, 1988). The crystallographic structure of the zinc finger domains reveals that each finger interacts with a triplet within the DNA substrate (Kim *et al.*, 1996). The authors hypothesised that the modular structure of the zinc finger domains could be used to engineer proteins recognising any DNA target sequence. Whereas theoretically all possible triplet recognition sequences could be designed by substituting individual amino acids, in practice context-dependent effects between adjacent fingers restrict the possibilities. As a result, only GNN and ANN, and a few CNN and TNN ZFN have been successfully constructed (Durai *et al.*, 2005). Programs have been developed for the rational design of ZFNs taking into account these restrictions, *e.g.* CoDA (Sander *et al.*,

2011). With the current archive of CoDA units, a potential ZFN target site can be found approximately once in every 400-500 bp of random genomic sequence.

Linked with an endonuclease, the hybrid ZFN could bind to a target sequence and induce a DSB, which would be repaired by any of the existing cellular DNA repair mechanisms. The bacterial type IIS restriction endonuclease *Fok*I has been specifically used for this purpose as it functions only as a dimer (Bitinaite *et al.*, 1998), which further increases the specificity of the ZFN action. Two ZFNs therefore need to bind to opposite strands of the DNA helix to be functional. The DNA binding domain of the original *Fok*I has been replaced by the zinc finger binding domains, creating chimeric ZFNs (Kim *et al.*, 1996).

Inducible modifications: Small insertions or deletions, or gene correction, replacement or insertion, depending on the provision of a donor template (see Section 5.1.2). In yeast, introduction of a ZFN in the presence of a template DNA frequently results in repair by homologous recombination, leading to gene addition or replacement, but this repair pathway is less common in other organisms. In rice, 1 to 11 bp deletions and substitutions were found in the ZFN-targeted *SSIVa* starch biosynthesis gene in four plants, resulting in premature stop codons and gene knock-down (Jung *et al.*, 2018).

Result: Without co-delivery of a donor template, the use of a ZFN may result in a gene knockout through the generated codon frameshift (or creation of a stop codon) as a result of the random indels created at the target site. Larger deletions can be obtained by using two ZFN pairs targeting sequences spaced around the region to be deleted. When a donor template with homology arms is co-delivered with the ZFN correction of a malfunctioning gene, or replacement of the endogenous gene by a desired coding sequence are possible. In addition, insertion of a new sequence (cis/intragenesis or transgenesis) may be achieved.

Target organisms: The ZFN technique has been used in yeast and other fungi, plants (maize, wheat, tobacco, rapeseed, rice; reviewed in Zaman *et al.*, 2019), animals (Drosophila, roundworm, zebrafish, rats, pigs, human), mostly for research purposes or for production of knockout model animals derived from treated embryonic cells (like rats and pigs) (Hauschild *et al.*, 2011). Some larger animals have been edited with ZFN for increased meat production (Qian *et al.*, 2015; Jiang and Shen, 2019) or disease resistance (Lillico *et al.*, 2016).

Known off-target effects: In some cases, cytotoxicity of ZFNs has been observed, which may be related to off-target cleavage of the DNA. Several approaches have been successfully described to reduce such effect, including making them more specific by use of more zinc fingers (4 to 6 per monomer) or using preferentially heterodimerising ZFNs (Miller *et al.*, 2007). Other strategies to lower off-target activity include decreasing their binding affinity (Pattanayak *et al.*, 2011), or lowering ZFN expression, *e.g.* through directly delivering ZFN mRNA or protein to the cells instead of the DNA expressing the proteins inside the cells (Bilichak *et al.*, 2020). In plants, limited successes have been obtained with ZFN technologies due to the typically low rate of HR in plants (reviewed in Weinthal *et al.*, 2010). In one successful study with maize, no off-target effects were identified in the potential ZFN off-target sites of five plants resulting from a sequence insertion in a phytate biosynthesis gene (Shukla *et al.*, 2009). Also in pig fibroblast cell editing, a nuclease assay showed that the ten most likely off-target cleavage sites were not modified (Hauschild *et al.*, 2011).

Limitations and gaps in knowledge: ZFNs suffer from target site availability, activity, and occasionally toxicity (Sander *et al.*, 2011). Engineering a new ZFN for a desired target sequence is not as straightforward as initially thought and this, together with the potential off-target cleavage activity of ZFNs, has limited their widespread use for genome alteration to a number of reported applications (Hauschild *et al.*, 2011; Bilichak *et al.*, 2020).

5.1.1.3 Transcription activator-like effector nucleases

Introduction: Transcription activator-like effectors (TALE), identified in several bacteria of the *Xanthomonas* genus, have been fused to endonucleases, mostly *Fok*I, to create an artificial site-specific transcription activator-like effector nuclease (TALEN) with modular flexibility (Christian *et al.*, 2010). Upon infection of plants with bacterial pathogens of the *Xanthomonas* genus, TALE proteins enter the cell nucleus, bind effector-specific DNA sequences, and transcriptionally activate gene expression, thereby increasing susceptibility of the host. Whereas the widespread adoption of ZFNs and meganucleases has been hindered by the challenge in engineering new DNA binding specificities, TALENs are much easier to predictably design and assemble. The DNA binding properties of these rather large proteins are characterised by a high specificity, tolerating a minimal number of mismatches in the target sequence (Juillerat *et al.*, 2014).

Mechanism: Like ZFNs, TALENs are engineered proteins consisting of a DNA binding domain formed by in tandem arrayed TALE proteins linked to the catalytic domain of a DNA endonuclease, usually *Fok*I. Each of the TALE proteins recognises a single base of a DNA sequence and usually 10-21 individual TALE are combined

and linked to *Fok*I. Each chimeric TALEN acts as a dimer due to the dimerisation requirement of *Fok*I. DNA cleavage by *Fok*I generates 5' overhangs (Smith *et al.*, 2000).

Compared to ZFNs, TALENs possess a broader targeting range and are less difficult to engineer (Boch *et al.*, 2009; Čermák *et al.*, 2011). Moreover, TALENs appear to be more mutagenic (*i.e.* more efficient) than ZFNs and are highly specific (Juillerat *et al.*, 2014). Like other SDNs, TALEN may be applied in the absence or presence of a donor template (Lee *et al.*, 2014). When two neighbouring TALEN sites are targeted, deletion of the intervening sequence has been achieved in animals (Carlson *et al.*, 2012) and in plants (Shan *et al.*,2013).

In animal cells, TALENs may be delivered *e.g.* through transfection of plasmid DNA (Beurdeley *et al.*, 2013) or by cytoplasmic injection of TALEN mRNAs into embryos (Carlson *et al.*, 2012). Viral delivery is hindered by the large size of the TALEN vector constructs (Holkers *et al.*, 2013; Ain *et al.*, 2015). In plants, stable insertion of the TALEN construct is often required (*e.g.* with *Agrobacterium* or ballistic bombardment), and the transgenes are subsequently segregated away in subsequent generations to obtain transgene-free plants (Haun *et al.*, 2014; Wang *et al.*, 2014; Li *et al.*, 2012 & 2016b). In potato TALEN-induced genome modifications were obtained by transient expression of plasmids in protoplasts (Clasen *et al.*, 2016).

Transcription activator-like effector nucleases (TALEN): detailed mode of action

The sequence specificity of the family of TALE proteins is driven by a domain composed of repeated motifs of 33–35 amino acids. These motifs are tandemly repeated as many as 30 times and have a largely invariant sequence. The specificity results from two centrally located polymorphic amino acids, the so-called repeat variable diresidues (RVDs), located at positions 12 and 13 of a repeated unit. Each single base of a DNA target is contacted by a single repeated unit in a 5'-3' direction, whereby the amino acid at position 13 contacts, in the major groove, the top DNA strand base, and the amino acid at position 12 participates in the stabilisation of the repeated units (Juillerat *et al.*, 2014). After deciphering of the one-to-one RVD/nucleotide association code (NI:A, HD:C, NN:G and NG:T), TALE DNA binding arrays specific for any DNA sequence could be created (Boch *et al.*, 2009).

Christian *et al.* (2010), relying on the known design of ZFNs, fused the TALE DNA binding function to the catalytic domain of a bacterial endonuclease that was previously used for ZFNs, *i.e. Fok*I, to create the chimeric genome editing tool TALEN. Both native and custom-designed TALE-nuclease fusions were shown to generate DNA double-strand breaks at specific, targeted sites. Such TALENs act as a dimer due to the *Fok*I properties and require engineering and simultaneous expression of two TALENs to function. When a new TALEN target is chosen, each site should start with a T in the DNA and have at least an additional 10–12 bp, although 15–21 bp are more common. The binding sites are in opposite orientation and are separated by 12–20 bp due to the additional protein sequence between the TALE modules and the nuclease domain (Carroll, 2014).

One approach to reduce the large size of the TALEN monomers (>100 kDa, >1800 amino acids), has been the development of a single-chain TALEN (scTALEN), in which two *Fok*I nuclease domains are fused on a single polypeptide with a 95 bp polypeptide linker (Sun and Zhao, 2014). This also makes it easier to design new TALENs as only one binding site has to be chosen instead of two. In another approach, the TALE proteins were fused to the cleavage domain of a homing endonuclease, I-*Tev*I (Beurdeley *et al.*, 2013). Moreover, they found that the amino-terminal *Tev*I fusions function as natural cleavases, while the carboxy-terminal fusions function as natural nickases. Nickases generate single-strand breaks instead of DSBs, potentially favouring homologous recombination instead of the error-prone NHEJ (Fauser *et al.*, 2014). Other improvements, *e.g.* in the DNA binding domain of TALENs, have been suggested (reviewed in Sprink *et al.*, 2015).

Inducible modifications: Short deletions (rarely short insertions) when using a single TALEN pair, larger deletions (sometimes also inversions) when using two TALEN pairs aimed at gene knockouts; in the presence of a donor template, also gene correction, replacement or insertion have been reported.

Shan *et al.* (2013) successfully targeted seven genes in *Brachypodium* and four in rice by specific TALEN pairs. The resulting mutations were predominantly short deletions of 1–20 bp, indicating DNA repair by the NHEJ pathway. Larger regions of DNA were deleted (or inverted) in pig cells through application of two TALEN pairs with specific cutting sites several thousand nucleotides apart (Carlson *et al.*, 2012). Using the same strategy, a 1.3 kb deletion and one case of a sequence inversion was observed in rice (Shan *et al.*, 2013). Haun *et al.* (2014) used a TALEN pair targeting a sequence conserved in two fatty acid desaturase genes (*fad*2-1A and *fad*2-1B) and obtained deletions of 14-63 bp in both genes. Wang *et al.* (2014) obtained short deletions of 1-10 bp in three powdery mildew resistance homoeoalleles in hexaploid bread wheat using a single TALEN pair recognising two conserved sequences separated by an 18 bp spacer region.

Homologous recombination following co-delivery of TALENs and a donor template has been confirmed in cattle, creating an 11.5 kb insertion of a human serum albumin gene and replacement of the endogenous BSA sequence (Moghaddassi *et al.*, 2014). In rice, ballistic delivery (with a gene gun) of a TALEN expression construct and donor DNA resulted in a successful sequence replacement, introducing the donor herbicide resistance (acetolactate synthase, *als*) gene containing two-point mutations compared to the endogenous *als* sequence (Li *et al.*, 2016b).

Result: Without co-delivery of a donor template, use of a TALEN may result in a gene knockout through the generated codon frameshift (or creation of a stop codon) as a result of the random indels created at the target site. Indels do not necessarily have to result in gene knockouts, as demonstrated in rice: deletion of the effector-binding element for the AvrXa7 pathogen from the promoter of a virulence susceptibility gene resulted in resistance against bacterial blight without compromising the endogenous function of the gene (Li *et al.*, 2012).

Larger deletions can be obtained by using two TALENs targeting sequences spaced around the region to be deleted. When a donor template with homology arms is co-delivered with the TALENs correction of a malfunctioning gene, or replacement of the endogenous gene by a desired coding sequence are possible. In addition, insertion of a new sequence (cis/intragenesis or transgenesis) may be achieved.

Target organisms: TALENs have been applied mainly in animals and plants (Carroll, 2014).

Known off-target effects: TALENs are reported to be more specific compared to ZFNs, with a low tolerance to target mismatches (Zhang *et al.*, 2013b). Specificity is further enhanced by improvements similar as for ZFNs, *e.g.* use of obligatory heterodimerising *Fok*I domains (Carroll, 2014). In cattle, whole genome sequencing of the progeny of TALEN-edited hornless bull confirmed the absence of other modifications in the genome (Young *et al.*, 2020). No off-target modifications have been described in the consulted literature; however, this is based on a limited number of papers available.

Limitations and gaps in knowledge: Both ZFN and TALEN technologies require engineering of DNA-binding domains for individual targeting applications, demanding significant effort and requiring some empirical testing. Delivery of the large-sized TALEN constructs also hinders several viral delivery systems for animal cells (Holkers *et al.*, 2013; Ain *et al.*, 2015).

5.1.1.4 Clustered Regularly Interspaced Short Palindromic Repeats-associated protein (CRISPR-Cas)

Introduction: The CRISPR-Cas system is the most recently developed genome editing tool and it is widely considered the most powerful and most versatile one among the SDNs to edit specific sites of the genome. The technique was first applied for genome editing in 2012-2013 (the discovery milestones are summarised in Mojica and Montoliu, 2016) and it has since seen an almost exponential increase in further improvements and in number of applications in research and development, as evidenced from the literature (Huang *et al.*, 2019a). Its ease of use, high efficiency and enhanced flexibility with the possibility of multiplexing has revolutionised plant and animal research and breeding (Schindele *et al.*, 2020; Wu *et al.*, 2018) and offered new opportunities for human gene therapies (Broeders *et al.*, 2020), diagnostic analysis (Myhrvold *et al.*, 2018), industrial biotechnology (Zhang *et al.*, 2020b) and even biosensing applications (Li *et al.*, 2019a).

The CRISPR/Cas technique originates from type II prokaryotic adaptive immunity systems that help protecting bacteria and archaea against invading phages and plasmids (Barrangou *et al.*, 2007). These organisms 'remember' the sequences of previously invading viral genomes and protect themselves by recognising and cutting those sequences when they are encountered again (recent evidence supports the existence of comparable viral defence systems in eukaryotic genomes, see Ophinni *et al.*, 2019). The prokaryotic CRISPR locus harbours up to a few hundred short stretches (30-40 bp) of the invading foreign nucleic acids, so-called protospacers, separated by largely invariable and partially palindromic repeats of 25-35 bp each. These protospacers are transcribed into CRISPR RNA (crRNA). A second RNA known as the trans-activating CRISPR RNA (tracrRNA) is transcribed from a genomic locus upstream of the CRISPR locus and forms a complex with the crRNA. The crRNA:tracrRNA complex associates with a Cas protein (the nuclease) and creates an active ribonucleoprotein (RNP) complex that targets foreign nucleic acids for degradation (Mojica and Montoliu, 2016; Maeder and Gersbach, 2016).

The prokaryotic CRISPR/Cas system has been engineered for genome editing applications by fusing the tracrRNA and crRNA into a single-guide RNA (sgRNA) which binds complementarily to a DNA target. Researchers can design the guide RNA to target nearly any place in the genome allowing the Cas nuclease to make a double-strand break (DSB) at any targeted site (Jinek *et al.*, 2012).

Technological improvements and variations in CRISPR-Cas genome editing strategies are being reported in the recent literature with unsurpassed speed (Huang *et al.*, 2019a). In this section the original CRISPR/Cas technique will be described, which results in a DSB in the genome. This is exemplified by the most common CRISPR tools derived from the CRISPR-Cas9 system found in *Streptococcus pyogenes*, known as SpyCas9. Other Cas proteins have since been discovered in this species and in other bacteria, some of which revealed distinct functionalities (Koonin *et al.*, 2017; Shmakov *et al.*, 2017). Variations of this basic technique, *e.g.* use of catalytically inactivated Cas proteins inducing only a single-strand break or no DNA break at all, which are often fused with other effector proteins, will be discussed in subsequent chapters (Group 2 and Group 3 NGTs). Also Cas proteins acting on RNA instead of DNA will be presented in a separate chapter (Group 4 NGTs).

Mechanism: The essential components of the basic CRISPR/Cas tool are the Cas endonuclease and a single-guide RNA (sgRNA) designed to recognise the target sequence. The target sequence must contain a protospacer adjacent motif (PAM), which is usually an absolute prerequisite for the Cas protein to induce a DSB (Sternberg *et al.*, 2014). The recognition, binding and cleavage processes occur in a multistep mechanism. Cas-sgRNA complexes start searching for target sequences through recognition of PAM sequences. The Cas protein directly binds the PAM sequence through protein-DNA interactions and subsequently unzips the downstream DNA sequence. The Cas protein then interrogates the extent of base pairing between one strand of the DNA target and the sgRNA. Sufficient complementarity between the two drives target cleavage, otherwise the Cas-sgRNA complex rapidly dissociates from the DNA (Sternberg *et al.*, 2014). The widely used Cas9 from *Streptococcus pyogenes* (SpyCas9) recognises an NGG PAM, and to a lesser extent, an NAG PAM. Other Cas proteins recognise other PAM sequences, sometimes up to several nucleotides long. Extensive engineering of various Cas proteins has been achieved for the recognition of alternative PAM sites (Leenay and Beise, 2017; Hu *et al.*, 2018). This has recently led to the creation of a near PAMless SpyCas9 recognising almost any DNA triplet sequence, thus widening the range of genome editing to any target (Walton *et al.*, 2020).

Whereas the PAM sequence is an initial anchoring point, the specificity of genome editing using CRISPR-Cas is determined by the sgRNA, *i.e.* by RNA-DNA interactions rather than by protein-DNA binding as for the other SDNs described above. The sgRNA is a chimeric RNA sequence approximately 100 nucleotides long with a target DNA recognition region at the 5′ end of approximately 20 bp long. This DNA binding region is followed by an invariable hairpin structure through fusion of the 3′ end of crRNA to the 5′ end of tracrRNA, retaining the base-pairing interactions that occur between the natural tracrRNA and the crRNA molecules (Jinek *et al.*, 2012).

With a properly designed guide sequence, on-target genome editing by CRISPR-Cas9 is usually efficient. However, initial CRISPR-Cas9 tools developed for genome editing were sometimes hampered by off-target effects (Lin *et al.*, 2014). Non-specific binding of the Cas9 (and other Cas proteins) has been thoroughly investigated and improvements in specificity of the technique have been achieved either by engineering the Cas nuclease, regulating its expression, or through optimisation of sgRNA design. Several sgRNA modifications for reduced off-target activity have been explored in different systems which include truncations, extensions, insertion of DNA or LNA (locked nucleic acids) nucleotides, addition of a 5'-end hairpin structure and others (reviewed in Wu and Yin, 2019). The so-called seed region of the sgRNA, *i.e.* the PAM-proximal 10-12 nucleotides at the 3' end of the 20 bp recognition region, is of particular importance for specificity. Mismatches in this seed region severely impair or completely abrogate target DNA binding and cleavage, whereas close sequence similarity in the seed region often leads to off-target binding events even with many mismatches elsewhere (Pattanayak *et al.*, 2013).

For genome editing applications, the coding sequences of Cas9 and sgRNA are cloned into a single or two separate expression vectors and delivered to the cells to be edited. Modular cloning systems, *e.g.* Golden Gate, have been optimised for CRISPR-Cas mediated genome editing (Vad-Nielsen *et al.*, 2016). These vectors can be delivered to the cells by various means and they affect stable or transient expression of the Cas and sgRNA sequences. An alternative to DNA delivery is the use of the mRNA or pre-assembled Cas9-sgRNA ribonucleoprotein (RNP) to affect genome editing (Ran *et al.*, 2017). This reduced the frequency of off-target effects, as has been shown for several plants and animals (Woo *et al.*, 2015; Liang *et al.*, 2015; Kim *et al.*, 2017a; Liang *et al.*, 2017a; Malnoy *et al.*, 2016; Alok *et al.*, 2020).

CRISPR-Cas: detailed mode of action

The SpyCas9 protein is a large monomer of 1,368-amino-acids (Jiang and Doudna, 2017) consisting of two functional nuclease domains (HNH and RuvC), in addition to other functionalities in crRNA maturation and spacer acquisition. In the absence of a sgRNA the Cas9 protein is kept in an inactive configuration incapable of DNA binding, but upon sgRNA loading the protein structurally changes into an active DNA surveillance complex capable of DNA cleavage. The Cas9 and similar Cas proteins are therefore called RNA-guided endonucleases. The HNH-like nuclease domain (HNH is a class of homing endonucleases, whereby H and N refer to histidine and asparagine, respectively) cleaves the DNA strand complementary to the guide RNA sequence (target strand) and the RuvC-like nuclease domain is responsible for cleaving the DNA strand opposite the complementary strand (non-target strand). The Cas9 RuvC nuclease domain, composed of three RuvC motifs, shares structural similarity with the retroviral integrase superfamily members characterised by an RNase H fold. In contrast, the HNH nuclease domain adopts the signature ββα-metal fold shared with other HNH endonucleases. These nuclease domains need respectively two- and one metal ion for catalysis. Structurally, the nuclease domains are located in one lobe distinct from the second lobe of the protein harbouring the recognition functions of Cas9 and both lobes connect through two linker regions. The C-terminal part of the protein is responsible for PAM recognition. This structure is consistent with Cas9 mutagenesis studies showing that mutating either the HNH (H840A) or the RuvC domain (D10A) converts Cas9 into a nickase (nCas9), whereas mutating both nuclease domains of Cas9 (so-called 'dead or deactivated Cas9' or dCas9) leaves its RNA-guided DNA binding ability intact while abolishing endonuclease activity (Jinek *et al.*, 2012). The nickase and dCas9 variants are described separately in the chapter on Group 2 NGTs.

Upon sgRNA loading, the activated Cas9-sgRNA complex searches the genome for target sites with the appropriate PAM sequence. Once a PAM is recognised, Cas9 triggers local DNA melting at the PAM-adjacent nucleation site, followed by RNA strand invasion to form an RNA-DNA hybrid and a displaced DNA strand (termed R-loop). The Cas9 enzyme then undergoes a further conformational re-arrangement that positions the HNH nuclease domain for cleavage of the target DNA strand. The conformational change of HNH simultaneously results in a large conformational change of the loop linkers, which in turn directs the non-target strand to the RuvC catalytic centre for concerted cleavage (Jiang and Doudna, 2017). Both nuclease domains cleave the DNA at a specific site 3 bp from the NGG PAM sequence to produce a predominantly blunt-ended DSB.

Given the mechanism of Cas9 DNA interaction it is not surprising that Cas activity depends on DNA target accessibility, which is largely affected by chromatin structure. One modification of this genome editing tool aimed at opening up the chromatin is the use of a second catalytically-inactive Cas9 (called proxy-Cas9) to bind the DNA in the vicinity of the target site (Chen *et al.*, 2017). Activity can also be enhanced by coupling the Cas9 with chromatin modulating peptides (CRISPR-Chrom) (Ding *et al.*, 2019). A different strategy for enhancing Cas activity while reducing off-target effects is the engineering of its amino acid sequence to generate high-fidelity Cas versions (Kleinstiver *et al.*, 2016; Slaymaker *et al.*, 2016; Casini *et al.*, 2018; Hu *et al.*, 2018).

A wide number of distinct Cas proteins have been discovered over the past few years (Barrangou, 2015). Cas proteins are commonly classified in two classes and six subgroups, although several additional subtypes have been suggested (Shmakov *et al.*, 2017). The more common Class 1 systems involve effector complexes formed by multiple (4-7) subunits, whereas in Class 2 systems, single multi-domain proteins constitute the effector complexes. Furthermore, based on their effector architectures each class has been divided into several subtypes (Class 1: types I, III and IV, and Class 2: types II, V and VI) with unique signature proteins (Koonin and Makarova, 2019).

- Type I systems of the **Class 1** category are the most common type of CRISPR-Cas systems in nature, comprising a multimeric DNA-targeting complex termed 'Cascade' and the endonuclease Cas3 (Pickar-Oliver and Gersback, 2019). Cas3, with its unique nickase and helicase activity, has been used as antimicrobial tool (Gomaa *et al.*, 2014).
- **Class 2** proteins, particularly Cas9 and Cas12a, have been mostly used for genome editing due to their monomeric structure. Orthogonal Cas9 proteins from different species exhibit limited sequence similarity and highly variable length (~900–1,600 amino acid residues), aside from the conserved HNH and RuvC nuclease domains that are required for dsDNA cleavage (Karvelis *et al.*, 2017). Among them, subtype II-A Cas9 from *S. pyogenes* (1,368 amino acids) is the most studied and most commonly used Cas9 version for genome engineering, while Cas9 orthologs from *Staphylococcus*

aureus (SaCas9, subtype II-A) and *Neisseria meningitides* (NmeCas9, subtype II-C) are much smaller (<1,100 amino acids) and more in line with viral delivery systems (Jiang and Doudna, 2017). Li *et al.* (2019c) reports a list of various Cas9 variants with improved characteristics for genome editing.

- Unlike Cas9, the Cas12a (formerly called Cpf1) system does not require a tracrRNA to mature the crRNA and to form an effector complex for its cleavage activity. Instead, it only requires a single short ~40 nucleotide CRISPR RNA (crRNA) to program target specificity (Kleinstiver *et al.*, 2019). Additionally, Cas12a possesses RNase activity, enabling poly-crRNA transcript processing for multiplex targeting applications (Zetsche *et al.*, 2017). It has also been reported to be less cytotoxic at high expression levels compared to Cas9 (Wendt *et al.*, 2016). After assembly, the Cas12a effector complex recognises a TTTV PAM for the initiation of binding and interrogation of target sites. Its seed sequence was illustrated to range from 1 to 10 bp proximal to the PAM. The Cas12a protein has a Nuc nuclease domain, distinct from the HNH domain of Cas9, for cleaving the target strand and a RuvC domain that cleaves the non-targeted strand. The nuclease domains cut the target DNA at bp 18 on the non-targeted strand and bp 23 on the targeted strand distal to the PAM, generating 5'-overhang ends (Jiang and Doudna, 2017; Paul and Montoya, 2020). Enhanced variants of Cas12a have also been engineered displaying a wider targeting range and improved on-target activity and fidelity (Kleinstiver *et al.*, 2019). Recently, Karvelis *et al.* (2020) discovered 10 exceptionally compact (422-603 amino acids) CRISPR-Cas12f nucleases capable of dsDNA cleavage in a PAM-dependent manner.
- A different subtype (VI) of Cas effector proteins is Cas13, which possesses RNA-guided RNase activity, targeting RNA instead of DNA, and has been used as diagnostic tool (Ackerman *et al.*, 2020) or for RNA editing (Cox *et al.*, 2017; see Group 4 NGTs for RNA editing).
- A new Class 2 system (type V) encoding miniature effectors of 400-700 amino acids, coded Cas14(a to h), have been identified, acting as PAM-independent single-strand DNA (ssDNA) nucleases (Harrington *et al.*, 2018). Due to its small size and ssDNA preference Cas14 variants may have a role in resistance against ssDNA viruses. A Cas14-DETECTR tool, based on isothermal amplification, has been developed for SNP genotyping (Harrington *et al.*, 2018).

Besides these known Cas enzymes, many more CRISPR-Cas systems are to be identified in the large bacterial and archaeal lineages in the coming years.

Inducible modifications: In the original CRISPR-Cas system a DSB is generated, which is most frequently repaired through NHEJ, thereby occasionally creating small mutations in the form of single nucleotide variations, or short nucleotide insertions or deletions (up to a few nucleotides). Using a single sgRNA, deletions up to 600 bp have been found in a large genome editing study in mice, but most target site alterations were considerably shorter (median of 9 bp) and insertions were also occasionally identified (Shin *et al.*, 2017). For gene knockouts, often two sgRNAs for the same gene are administered to the cells, resulting in two DSBs and deletion of the intervening sequence (which can be up to several thousand bp long). In the presence of a donor template HR may occur resulting in gene correction, replacement or addition (discussed in the following section). One important feature of the CRISPR-Cas technique is its easy adaptability to multiplex editing, based on the delivery of multiple sgRNAs to the cells. There are various strategies for the expression of multiple sgRNAs by one promoter as a single transcript that is further processed inside the cells to release individual sgRNAs (Vats *et al.*, 2019). Multiplex editing allows the functional removal of a complete metabolic pathway or several alleles or genes at once (Ma *et al.*, 2015; Wang *et al.*, 2017; Zhou *et al.*, 2019b; Zhang *et al.*, 2019; Wolter *et al.*, 2019). Besides for genome editing, Cas9 has been used for inducing viral resistance in transgenic plants and animals through expression of Cas9 together with virus-targeting sgRNAs (Price *et al.*, 2015; Niu *et al.*, 2017; Zhang *et al.*, 2018).

Result: Without co-delivery of a donor template, use of the CRISPR-Cas system with one sgRNA may result in a gene knockout through the generated codon frameshift (or creation of a stop codon) as a result of the random indels created at the target site. Larger deletions can be obtained by using two sgRNAs targeting sequences spaced around the region to be deleted. When a donor template with homology arms is co-delivered with the CRISPR-Cas system correction of a malfunctioning gene, or replacement of the endogenous gene by a desired coding sequence are possible by one of several repair pathways (reviewed in Section 5.1.2). In addition, insertion of a new sequence (cis/intragenesis or transgenesis) may be achieved.

The flexibility of the CRISPR-Cas system makes multiplexing possible, targeting several genes at the same time, whereby the alterations generated may be homozygous (all alleles at a targeted locus having the same alteration, *e.g.* a 2 bp deletion), bi-allelic (each allele having a different alteration) or heterozygous (the

sequence of only one allele is altered) (Zhang *et al.*, 2014b; Najera *et al.*, 2019). A complete gene family or metabolic pathway may be modified in one step (Wang *et al.*, 2017; Najera *et al.*, 2019). Multiplexing also makes the *de novo* domestication of desirable traits from wild races possible in much shortened timeframes compared to conventional breeding (Østerberg *et al.*, 2017; Zsögön *et al.*, 2018; Hickey *et al.*, 2019; Lacchini *et al.*, 2020; Courtier-Orgogozo and Martin, 2020).

Target organisms: The CRISPR-Cas genome editing technique can be used in various plants (Sedeek *et al.*, 2019; Schindele *et al.*, 2020), animals including mammals (Lee *et al.*, 2020) and insects (Gantz and Akbari, 2018), fungi and yeast (Morio *et al.*, 2020; Hassan *et al.*, 2020), and other microorganisms (Ren *et al.*, 2020).

In plants, CRISPR-Cas genome editing has been successfully applied to major crops like maize, wheat, soybean, rice, cotton, potato and many other plants, including fruit trees (Kim *et al.*, 2017a; Malnoy *et al.*, 2016; Liang *et al.*, 2017a; Woo *et al.*, 2015).

In animals, the CRISPR-Cas system may be delivered to somatic cells, which may subsequently be used for somatic cell nuclear transfer, or it can be directly injected into embryonic cells (Lee *et al.*, 2020).

Genome editing in insects was initially applied to model species like *Drosophila* fruit flies and later to various other insects (Reid and O'Brochta, 2016). Examples of applications include knockouts of up to several thousand kb with application of two sgRNAs, or knockins of up to 17.3 kb in mosquitoes as part of a gene drive approach to combat the spread of vector diseases (Gantz and Akbari, 2018).

In highly polyploid genomes like those of some cyanobacteria, CRISPR-Cas offers a considerable time-saving opportunity to produce genome edits through HR at all homozygous loci (Wendt *et al.*, 2016; Behler *et al.*, 2018).

Known off-target effects: Three major types of off-target regions have been described, including those with substitutions or mismatches compared to the target region (particularly in the non-seed region), those with insertions and/or deletions (indels) in comparison with target DNA or sgRNA spacer (which may result in a small bulge of unpaired nucleotides), and those with a different PAM sequence (Manghwar *et al.*, 2020). A fourth type may be the unexpected off-targets in genomic regions which are not related to the target, such as initially reported for mice, but later contested (Montoliu and Whitelaw, 2018). In plants, off-target modifications may not necessarily result in a modified phenotype and may be segregated out in subsequent generations. Such effects may be much more critical for therapeutic and clinical applications (Zhang *et al.*, 2015b).

The identification of such off-target effects in initial CRISPR-Cas experiments has prompted investigations to mitigate or reduce such effects. Several approaches have been employed, including more careful target selection and sgRNA design, reducing the expression of Cas9 through use of weaker promoters, spatial or temporal control systems for Cas expression, introduction of Cas-mRNA or ribonucleoproteins (RNPs) instead of vector DNA, or use of high-fidelity Cas proteins (Wu and Yin, 2019; Hajiahmadi *et al.*, 2019; Gangopadhyay *et al.*, 2019; Broeders *et al.*, 2020; Manghwar *et al.*, 2020). Others have diverted from using the DSB-generating SDNs toward deactivated versions with partially or completely impaired nuclease functions (*e.g.* nickases or dCas), which have demonstrated to be less prone to off-target activity (Ran *et al.*, 2013). An interesting very recent addition to enhance specificity is the linking of a dCas9 to the obligate dimerising *Clo*51 nuclease, a proprietary technique called Cas-CLOVER reported to be void of off-target activity (https://www.genengnews.com/resources/webinars/cas-clover-the-clean-alternative-to-crispr-cas9/). Another recent development with potential for reducing off-target activity is the use of anti-CRISPR proteins found in bacteriophages to tailor Cas activity to specific cells or tissues (Hwang and Maxwell, 2019), or use of anti-CRISPR agents that can switch the CRISPR-Cas system on and off (Dolgin, 2020).

Limitations and gaps in knowledge: The CRISPR-Cas technology is currently still in full evolution and improvements and novel applications are regularly reported. At the same time, the technique is so diverse in its modalities that is should be considered a technology platform rather than a single technique.

It is obvious that a perfect knowledge of the genome sequence in a given individual organism is essential for performing genome editing with high precision and limited spurious activity. One of the topics that receives much attention in research is the balancing of Cas activity with the risk for off-target modifications, and this may need to be established and optimised for every single organism and cell type (Hajiahmadi *et al.*, 2019). Our knowledge on the importance of sequence variation at the target site for preventing off-target modifications is still evolving (Wang *et al.*, 2020a). Particular attention needs to be paid to the evaluation of off-target modifications both through careful bioinformatics analysis to assess the potential off-target sites (*e.g.* Cui *et al.*, 2020) and through experimental testing of the intactness of the genomic sequence at these

sites (Kempton and Qi, 2019). As for any NGT, delivery of the components to the cells or tissues, identification and selection of the modified cells and development of full organisms from them remain issues that may be more challenging for some applications compared to others.

5.1.2 Cellular repair pathways for DNA double-strand breaks

Double-strand breaks (DSB) in the DNA are natural phenomena in all living organisms, but they affect cellular functionality and disturb the normal development of the organism. All organisms have therefore developed a network of cellular signalling and repair mechanisms that detect and remove the lesions, thereby reconstituting the original genetic information (Manova and Gruszka, 2015). An intrinsic feature of some of these repair mechanisms is that they are not error-free and induce potentially transmissible mutational alterations. This creates genetic diversity and drives evolution (Fitzgerald and Roosenberg, 2019).

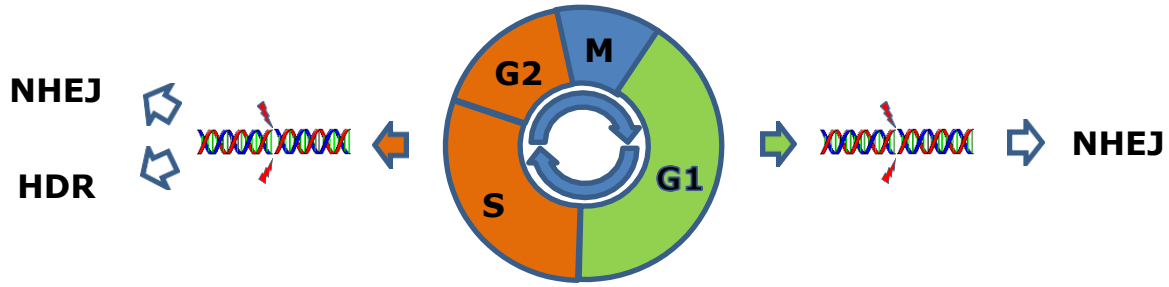
For genome editing, efficient DSB formation (Section 5.1.1) is the first step in a two-step process. The actual sequence editing is performed by the cell's DSB repair machinery. Presently, it is possible to dictate the precise site of genome editing by DSB formation but there is less control on the outcome of DNA repair. The choice of the repair pathway followed after a DSB is a balance between different repair options that is affected by multiple parameters and it has only now started to be understood. These parameters include the organism and the cell type, the cell cycle stage, the structure of the target chromatin and others (Her and Bunting, 2018).

DNA repair pathways in prokaryotes are based on the SOS regulatory network and are distinct from those in eukaryotes (reviewed in Maslowska *et al.*, 2019). The focus in this review is on eukaryotic responses to DNA damage.

Two major repair pathways operate in eukaryotic cells to repair DSBs, complemented by less common alternative pathways (Danner *et al.*, 2017). Both pathways are highly conserved from yeast to human (Lieber, 2010). As illustrated in Figure 3, a key determinant of repair pathway usage is the cell cycle phase (Her and Bunting, 2018). The most common repair pathway is **non-homologous end joining (NHEJ)** involving the direct resealing of the two broken DNA ends independently of significant sequence homology. Although being active throughout the cell cycle, NHEJ is relatively more important during the G1 phase. In contrast to NHEJ, **homology-directed repair (HDR)** requires a homologous DNA sequence to serve as a template for DNA-synthesis-dependent repair and involves extensive processing of the broken DNA ends. One form of HDR, homologous recombination, is extremely accurate, leading to precise repair of the damaged locus using DNA sequences homologous to the broken ends, often provided by the sister chromatids in dividing cells (Vítor *et al.*, 2020). Other HDR pathways are called microhomology-mediated end joining (MMEJ) and single-strand annealing (SSA). All HDR pathways are limited to the S and G2 phases of the cell cycle (Figure 3).

Figure 3. Cellular DNA repair choices operating during the cell cycle

Whereas DSBs are only repaired by the NHEJ pathway during the G1 interphase of the cell cycle, repair during the S and G2 phases may occur through either NHEJ or HDR. No homology-directed repair is occurring during the M phase. S: Synthesis (interphase), G1: Gap 1 (interphase), G2: Gap 2 (interphase), M: Mitosis (cell division).

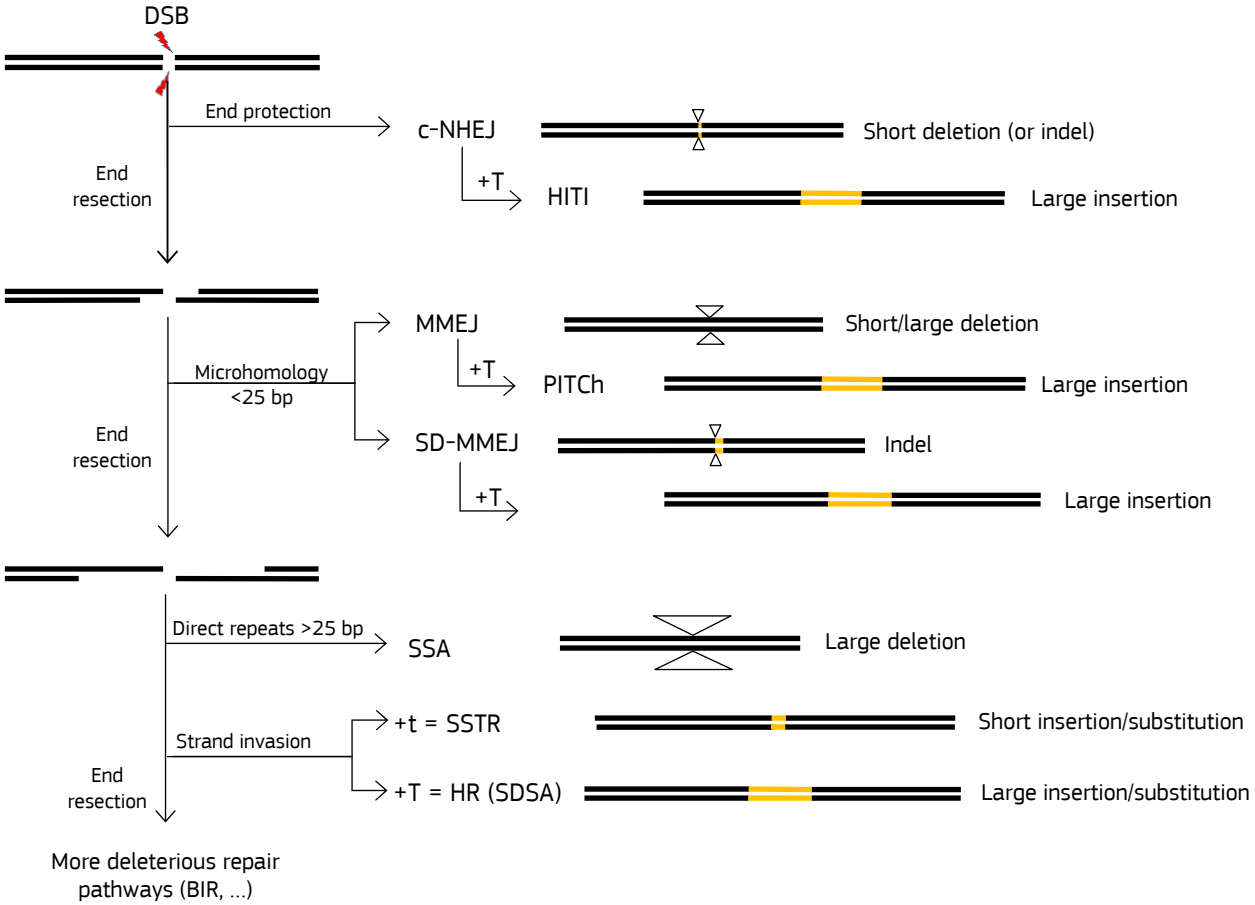


All pathways require at least some end processing by nucleases, utilisation of DNA polymerases, and a final ligation step to complete repair of the broken DNA (Pannunzio *et al.*, 2018). Figure 4 presents the basic mechanisms of the different repair pathways.

The initial phases are similar for both repair pathways and involve a number of proteins that make the chromatin more accessible to DNA repair factors (Danner *et al.*, 2017; Arnould and Legube, 2020). The first key step in the repair mechanism, *i.e.* whether or not DNA end resection occurs, determines which of the two major repair pathways will be used. Although NHEJ and HDR are the major mechanisms for DSB repair, alternative pathways have been discovered in recent years, mainly from work on mammalian cells (Her and Bunting, 2018; Danner *et al.*, 2017). All these repair mechanisms are described in more detail below. Finally, attempts to favour the less error-prone HR pathway over NHEJ are reviewed, based on experimental evidence from genome editing studies in several organisms.

Figure 4. Overview of the major cellular DNA repair pathways following a DSB and their use for genome editing

Abbreviations: c-NHEJ: canonical non-homologous end-joining; HITI: homology-independent targeted integration; MMEJ: microhomology-mediated end joining; PITCh: precise integration into target chromosome; SD-MMEJ: synthesis-dependent MMEJ; SSA: single-strand annealing; SSTR: single-stranded templated repair; HR: homologous recombination; SDSA: synthesis dependent strand annealing; BIR: break induced repair; +t, +T: addition of (short or long) donor template. Deletions and insertions are shown with arrowheads and orange boxes, respectively. Modified from various sources (Ceccaldi *et al.*, 2015; Rodgers and McVey, 2016; Danner *et al.*, 2017; Kostyrko *et al.*, 2017).



5.1.2.1 Non-homologous end joining (NHEJ)

Classical or canonical non-homologous DNA end-joining (c-NHEJ, here named NHEJ) is the predominant DSB repair pathway throughout the cell cycle, accounting for ~75 % of all natural DSB repair events, and for nearly all DSB repair outside of the S and G2 phases (Pannunzio *et al.*, 2018). Several lines of evidence indicate that NHEJ often acts first to attempt to repair DSBs. If NHEJ cannot be completed, then the DSB undergoes 'resection', in which one or both strands of the DNA duplex are degraded to produce a single-stranded DNA overhang suitable for alternative pathways of repair (Her and Bunting, 2018). Despite what its name suggests, micro-homology regions of between 1 and 4 bp are common during repair through NHEJ (Pannunzio *et al.*, 2018).

Further details on the non-homologous end joining repair pathway: Active within minutes after the generation of the DSB, NHEJ involves the direct ligation of the two DNA ends with simple end trimming. The first step in this process is the binding of the heterodimer Ku70/Ku80 to the broken DSB ends. Ku70/Ku80 bind to ends that have been protected from resectioning by accumulation of the mammalian checkpoint protein 53BP1 and the protein RIF1 (Danner *et al.*, 2017). The central role of Ku has been confirmed in plants and this repair pathway is, therefore, also named KU-dependent NHEJ (Shen *et al.*, 2017). Subsequently, the catalytic subunit of the DNA-dependent protein kinase (DNA-PKcs) is recruited to the broken ends and is activated to phosphorylate itself and its targets, including the nuclease Artemis, which trims DSB ends via its nuclease activity. Nucleotide addition can occur by the Pol X family polymerases, Pol μ and Pol λ. Ultimately, the ligase complex composed of DNA ligase IV, the scaffolding proteins X-ray cross-complementing protein 4 (XRCC4), XRCC4-like factor (XLF) and PAXX (paralog of XRCC4 and XLF), re-ligate the blunt DNA ends (Han and Huang, 2020). Many more proteins are involved in one or the other way in the NHEJ mechanism (Pannunzio *et al.*, 2018; Her and Bunting, 2018; Han and Huang, 2020).

Interestingly, after Ku binding to both ends of a DSB, the nuclease (N), polymerases (P), or the ligase complex (L) can bind in any order in a reiterative process (Lieber, 2010). This feature highlights the flexibility of NHEJ and explains the diversity of repair products generated for the same DSB configuration and DNA end sequence. In the simplest case, the ligase complex ligates the top strand, and then ligates the bottom strand, resulting in a fully repaired DSB site. This might occur at a DSB with two blunt DNA ends or two DNA ends with compatible overhangs. More typically, the nuclease or polymerases bind and modify either or both DNA ends. This typically involves removing short regions of the 5'- or 3'-overhangs by either exonucleolytic or endonucleolytic processing by Artemis to expose short regions of microhomology (1-4 nt) between the strands that can facilitate end joining. An alternative mechanism for generating such microhomology is template-independent nucleotide addition to one or both strands by the polymerases Pol μ or Pol λ. As long as the DSB is not fully repaired, nucleases can remove nucleotides from a DNA end, with Pol subsequently adding nucleotides to that very same DNA end. Similarly, XRCC4Lig4 can successfully ligate one DNA strand of a DSB whereas Artemis DNA-PKcs may reverse this by cleaving the newly ligated strand at the DNA gap generated by the ligation. Depending on the DNA end configuration, different proteins may be recruited to the break in an attempt to repair it. Two blunt DNA ends may only require Ku and XRCC4Lig4 for joining, whereas incompatible 3' ends may require processing by Artemis DNA-PKcs, and incompatible 5' ends may require XLF or PAXX for additional structural support. Using single molecule FRET (Förster resonance energy transfer) microscopy, the importance of the individual proteins in DSB repair and their interactions is becoming better understood (Zhao *et al.*, 2019). The flexible and reiterative NHEJ repair process forms the origin of the variation in indels identified following SDN-induced DSBs in the absence of a repair template (Pannunzio *et al.*, 2018). Alternatively, variation may result from the action of different repair pathways, including NHEJ and microhomology-mediated end joining (MMEJ), on the same DSB, as shown for Cas9-induced DSBs in human cells (Brinkman *et al.*, 2018) and *Arabidopsis* (Shen *et al.*, 2017).

Repair of a DSB by the error-prone NHEJ results in short deletions or (less frequently) insertions at the DSB site, usually of a few basepairs. This may result in frameshifts or nonsense mutations (Lieber, 2010). In a large study on 63 immunity-related genes in tomato edited with CRISPR-Cas9 and a single sgRNA, the modifications at the targeted sites were characterised in 245 T0 transgenic tomato plants (Zhang *et al.*, 2020a). Among those, 20 % were homozygous for the mutation, 30 % heterozygous, 32% had two different mutations (biallelic), and 18 % had multiple mutations (chimeric). The mutations were predominantly short insertions or deletions, with 87 % of the affected sequences being smaller than 10 bp. Deletions and insertions of 1 bp were predominant, but also larger deletions (up to 411 bp) and insertions (up to 50 bp) were identified in individual plants. The mutations from the T0 generation were stably transmitted to later generations, although new mutations were detected in some T1 plants. No mutations were detected in 18 potential off-target sites among 144 plants. A similar outcome was reported after CRISPR-Cas9 editing of three isoflavone synthase genes in soybean, although deletion size was only up to 162 bp at a single site and only 1 or 2 bp insertions were detected among 144 transformed hairy roots (Zhang *et al.*, 2019). Similarly, in 84 transgenic T0 maize plants edited with CRISPR-Cas9, deletions < 10 bp and insertions of 1 or 2 bp were prevalent in 18 targeted genes, the largest deletion was 136 bp, and the largest insertions (identified in 2 plants) were 10 and 11 bp (Doll *et al.*, 2019). In pig fibroblasts, a single ZFN generated 80 mutations in one gene, 90 % of which were indels <20 bp, with a preference for deletions (Qian *et al.*, 2015).

When two DSBs are generated in proximity to each other deletion, or sometimes duplication or inversion, of the intervening sequence is possible, leading to gene knock-outs. Also larger deletions can be targeted by using two sgRNAs, *e.g.* sequences up to 65 kb have been deleted in mouse cells with low efficiency using CRISPR-Cas9 (Zhang *et al.*, 2015a) and a fragment of 1.1 Mb was deleted in mouse zygotes (Mizuno *et al.*,

2015). In pig zygotes, delivery of two TALEN pairs directed against the same chromosome enabled the isolation of colonies harbouring either deletions or inversions of the intervening sequence (Carlson *et al.*, 2012). The sites targeted by the two SDNs may ligate in a perfect manner or reveal short indels indicative of the NHEJ repair mechanism acting on a DSB (Carlson *et al.*, 2012; Bonawicz *et al.*, 2019).

Occasionally, NHEJ has also been used to insert larger DNA sequences in the genome by a process called homology-independent template integration (HITI; Yamamoto and Gerbi, 2018). This repair mechanism, in contrast to HDR, has the advantage of being also active in non-dividing cells, which opens up opportunities in gene therapy (Suzuki and Izpisua Belmonte, 2018). In this process, a template DNA without homology arms is provided, designed to include Cas9 cleavage site(s) that flank the donor sequence. Cas9 therefore cleaves both the genomic target sequence and the donor plasmid, thereby generating blunt ends associated with both target and donor sequences. The linearised donor DNA plasmid can then be used for repair by the NHEJ pathway, allowing for its integration into the genomic DSB site, often error-free (Suzuki and Izpisua Belmonte, 2018). An alternative HITI method using ZFN heterodimers, called 'obligate ligation-gated recombination' (ObLiGaRe), has been applied to insert genes of several kb into target sites in various human cells and in flies with perfect precision in at least 75 % of the cases (reviewed in Yamamoto and Gerbi, 2018).

5.1.2.2 Homology-directed repair (HDR)

At least three other cellular DNA damage repair pathways operate in addition to the predominant NHEJ for resolving DSBs in the genome. All of these depend on some level of homology with a donor template, which can be provided by an endogenous or a delivered exogenous sequence. The first step in HDR is the end resection of the DNA ends created at the DSB in order to generate 3' single-strand (ss) DNA overhangs. Because of their reliance on ssDNA overhangs generated by end resection, these three pathways are restricted to the S or G2 phase (Ceccaldi *et al.*, 2016). DNA-end resection, which catalyses the nucleolytic degradation of broken ends in a 5'→3' direction, therefore, likely regulates pathway choice away from NHEJ (Han and Huang, 2020).

Further details on the homology-directed repair pathways: The key regulator 53BP1 is rapidly localised to DSB sites by recognising histone methylation and ubiquitination and blocks DNA-end resection, thereby promoting NHEJ. BRCA1 counteracts the end-protection activity of 53BP1 and its downstream factors to allow the activation of DNA-end resection in S/G2 cells followed by resection-dependent repair (Han and Huang, 2020). Both antagonistic proteins (53BP1 and BRCA1) are regulated in a complex manner which is not fully understood (Her and Bunting, 2018).

DNA-end resection follows a two-step process that consists of an initial resection, followed by extensive resection (Han and Huang, 2020). In the first step, the MRE11-RAD50-NBS1 (MRN) complex, which has both endonuclease and exonuclease activity, and associated CtIP protein initiate resection. MRE11 first cuts one DNA strand close to the break site, using its 5'–3' endonuclease activity, and then degrades the same strand using its processive 3'–5' exo-nuclease activity. Ku70/80 is displaced from the break site by MRE11-mediated end processing, thereby preventing further NHEJ activity. Several other proteins and post-translational modifications positively or negatively regulate the key proteins in this process (Ceccaldi *et al.*, 2016; Her and Bunting, 2018; Mejías-Navarro, 2020). The result of this first step in DNA-end resection is the creation of 3'-overhangs consisting of relatively few basepairs at the break ends, *i.e.* 20 bp in mammalian cells or 100–300 bp in yeast (Ceccaldi *et al.* 2016; Han and Huang, 2020). This makes the DSB ends available for the alternative non-homologous end joining repair pathway (MMEJ). In the second step, the short resected ends are further processed by BLM-DNA2 and EXO1 to generate long stretches of ssDNA, allowing the execution of homologous recombination (HR) or single-strand annealing (SSA) (Ceccaldi *et al.*, 2016). All these highly regulated events take place within a modified chromatin environment, which contributes to enhancing accessibility of the broken ends and recruitment of DNA repair factors (Vítor *et al.*, 2020). Although many contributing factors and regulation mechanisms have been elucidated in recent years, the triggers that determine DSB repair pathway choice following DNA-end resection are still not fully understood (Han and Huang, 2020).

5.1.2.2.1 Homologous recombination

In contrast to NHEJ, homologous recombination (HR) requires a homologous DNA sequence to serve as template for DNA-synthesis-dependent repair in the S- and G2-phases of the cell cycle. In dividing cells, this template is usually provided by the sister chromatid rather than the homologous chromosome, but also exogenously delivered DNA templates may act as donor sequence for HR. HR is extremely accurate, as it leads to precise repair of the damaged locus using DNA sequences homologous to the broken ends. Sequences at

the interior of the homologous arms may be inserted into the genome as well, which offers an appealing application for site-specific gene insertion. In contrast to end-joining, HR is mechanistically more complicated, involves a larger number of enzymes, and is thus comparatively slower but more accurate. The advance of genome editing technologies brought renewed interest in understanding the balance between the DSB repair pathways, as the inhibition of NHEJ repair promotes HR-based precise genome editing (Danner *et al.*, 2017; Brinkmann *et al.*, 2018). Most of our knowledge on the repair processes comes from studies on yeast and mammalian cells, but more or less comparable mechanisms have been discovered in plants thereafter (Vu *et al.*, 2019).

Further details on the homologous recombination repair pathway: Following extensive resection of the broken DNA ends, the newly created 3' ssDNA stretches, which may be >3 kb long in mitotic cells (Zakharyevich *et al.*, 2010), are initially protected from cleavage by the replication protein A (RPA) heterotrimer. BRCA1 subsequently recruits PALB2 and then BRCA2 to the break site, which can load the recombination mediator Rad51, an ortholog of RecA in prokaryotes (Her and Bunting, 2018). Rad51 family ATPases are key to HR, forming right-handed helical filaments on ssDNA that act as nucleoprotein scaffolds to direct their own and other proteins' activities, such as RAD52 (Wright *et al.*, 2018). The Rad51 filaments, which form microscopically visible foci in the nucleus, maintain the ssDNA in a B-form with triplets open for Watson–Crick pairing with complementary triplets in homologous duplex DNA (Danner *et al.*, 2017). The broken DSB end, now resected to ssDNA and assembled with the Rad51 filament and cofactors, then searches for a homology donor from which to initiate DNA synthesis. Microhomologies of as few as eight nucleotides promote extended lifetimes of Rad51-ssDNA-dsDNA complexes and result in the formation of a synaptic complex. Then the 3' end of the invading strand intertwines with its complement in the donor to form a primer- template junction competent for DNA synthesis, known as the displacement or D-loop. The basepairing between the ssDNA arms and the genomic DNA can be short or up to several hundred basepairs long (Wright *et al.*, 2018). Rad51 is replaced by the ATPase Rad54 and DNA synthesis by polymerase Pol δ progresses through the migrating bubble along the homologous region encountered. It is unclear what determines the end of DNA synthesis, but as the D-loop is disrupted there should be sufficient sequence homology to anneal to the second resected end of the DSB. Evidence accumulates that strand invasion and DNA synthesis within a D-loop is a reiterative process, possibly monitored by ssDNA translocases (Wright *et al.*, 2018). The newly formed ssDNA at the resected DNA end subsequently anneals with its complementary resected strand at the DSB, and primes DNA synthesis from the resected end to restore the original sequence. Alternatively, strand invasion may occur at both resected ends of the DSB, forming a double Holliday junction, followed by annealing of the two newly formed ssDNA strands and re-ligation of the ends (Mehta and Haber, 2014).

For genome editing, the HR pathway can be co-opted by providing an artificial DNA repair template containing sequence regions homologous to the DSB ends. The sequence between the homologous ends is then transferred into the targeted locus during HR, enabling gene correction by the generation of precisely modified 'knockin' alleles, or gene insertion (Table 1). Plasmid-based gene-targeting vectors with homology regions of >500 bp are suitable for introducing large sequence changes and insertions (Danner *et al.*, 2017).

Targeted gene replacement through HR has been achieved in plants already in 1988 (Paszkowski *et al.*, 1988), but occurs with low efficiency in most eukaryotic cells (Puchta, 2005; Pâques and Duchateau, 2007). HR is, however, very efficient in yeast and this organism has functioned as model for deciphering the HR mechanism (Aylon and Kupiec, 2004). Gene insertion through HR is strongly enhanced upon DSB generation by any of the SDNs. Using a single vector encoding both the TALEN and a 2 kb homologous sequence, Li *et al.* (2016b) obtained a frequency of HR in up to 6 % of rice transformants, introducing two point mutations in the acetolactate synthase (ALS) gene. Reporter gene insertion by HR in mice embryonic cells and zygotes using homology arms of less than 1 kb revealed a 27 % and 11 % efficiency, respectively (Zhang *et al.*, 2015a). To increase HR efficiency, both in plants and in animals, several strategies have been experimentally tested and will be reviewed in Section 5.1.2.3.

5.1.2.2.2 Microhomology-mediated end joining (MMEJ), an alternative end joining pathway

Another pathway for repairing DSBs is termed 'alternative NHEJ' (a-NHEJ, also called alt-EJ), which is an error-prone process that may start after partial end resection has occurred.

Further details on the microhomology-mediated end joining repair pathway: There are probably several distinct alt-NHEJ pathways (Simsek *et al.*, 2011), but the best studied is classified as microhomology-mediated end joining (MMEJ), requiring homology sequences of 5-25 nucleotides at the broken ends (Truong *et al.*, 2013). It involves the poly-ADP ribose polymerase PARP1, the MRN complex and its partner CtIP, whereas c-NHEJ uses the Ku70/80 heterodimer and DNA-PK catalytic subunit (Shen *et al.*, 2017). Once

resected ends successfully anneal via complementary base pairing, the flanking single-stranded regions are then subject to fill-in synthesis. A distinguishing feature of MMEJ in eukaryotes (except yeast) is the presence of nucleotide insertions at break sites following repair. These insertions are often derived from sequences close to the breaks and in some cases are copied from other chromosomes. Additionally, a number of nucleotides can be added *de novo* as a result of non-templated extension of the ends by Pol θ (Sfeir and Symington, 2015). Pol θ seems to act as a key regulator favouring MMEJ over other repair pathways by removing RPA from resected ends (Her and Bunting, 2018). For re-joining the (micro)homologous ends, ligase 3 (Lig3) is required for MMEJ, in contrast to c-NHEJ which is dependent on Lig4 (Simsek *et al.*, 2011; He *et al.*, 2015).

Using a MMEJ and HR competition repair substrate, MMEJ with short-range end resection was shown to be active at a substantial frequency to repair DSBs in mammalian cells even when both c-NHEJ and HR pathways are available (Truong *et al.*, 2013). This pathway has an important biological role in repairing DSBs in cells under genomic stress, but it usually generates longer deletions than c-NHEJ and it may also lead to chromosome translocations in mammals (Ghezraoui *et al.*, 2014). Depending on the orchestration of the repair near the DSB and the occurrence of microhomologies in its vicinity, MMEJ may have different outcomes, ranging from short or long deletions and/or insertions to chromosomal translocations that may be lethal (Ceccaldi *et al.*, 2016).

Another variant of MMEJ, termed synthesis-dependent MMEJ (SD-MMEJ), uses microhomologies synthesised *de novo* by a non-processive DNA polymerase and does not rely on significant pre-existing homology (Yu and McVey, 2010). This pathway was identified as the main mechanism driving the illegitimate genomic integration of foreign DNA in Chinese hamster cells (Kostyrko *et al.*, 2017).

MMEJ can be harnessed for targeted sequence insertion by providing DNA fragments which are flanked by short homology regions (5-25 bp) to the DSB ends (Sakuma *et al.*, 2016). Such microhomology regions can *e.g.* be chosen in the spacer region between the TALEN binding pairs (TAL-PITCh) or adjacent to the CRISPR-Cas induced DSB (CRIS-PITCh). This PITCh (Precise Integration into Target Chromosomes) approach for gene addition has been employed for the integration of large donor DNAs into target sites in human cells, silk worm embryos, frogs, and mice (Nakade *et al.*, 2014; Danner *et al.*, 2017; Yao *et al.*, 2017). In rice, CRISPR-Cas9 induced double DSBs in the same gene were repaired either by NHEJ or by MMEJ with comparable frequencies (Tan *et al.*, 2020).

5.1.2.2.3 Single-strand annealing (SSA)

Single-strand annealing is another repair pathway that is conserved among kingdoms and may start after significant end resection following a DSB (Bhargava *et al.*, 2016). SSA requires the presence of long homologous direct repeat sequences (26 to >500 bp) flanking the DSB (Zhang and Matlashewski, 2019).

Further details on the single-strand annealing repair pathway: The SSA pathway does not involve Rad51 binding or strand invasion such as HR. Instead, the DNA binding protein Rad52 plays a key role in the end joining between interspersed nucleotide repeats in the genome. Rad52 binds to the exposed repeats flanking the DSB and promotes their annealing (Ceccaldi *et al.*, 2016). Additionally, the remaining non-complimentary ssDNA sequences are nucleolytically removed by ERCC1 that associates with XPF to cleave 3′ ssDNA tails. The nuclease activity of the ERCC1/XPF complex is enhanced by Rad52. Following these steps of annealing and 3′ ssDNA tail removal, any gaps are filled by DNA polymerases and a DNA ligase connects the ends to complete DNA break repair (Bhargava *et al.*, 2016). Although this is homology-directed repair, one copy of the repeat and the intervening sequence between the repeats are deleted in the repair product, thus resulting in the loss of genetic information up to several thousand bp (Orel *et al.*, 2003; Puchta, 2005). In organisms with a large proportion of repetitive sequences, *e.g.* polyploids, the use of the mutagenic SSA pathway for gene repair is tightly controlled as it would otherwise lead to chromosomal rearrangements (Hu *et al.*, 2019).

As a result of the SSA mechanism of action described above, the outcome may be small or large deletions, but no new insertions of basepairs at the DSB site. The SSA pathway is the predominant DSB repair pathway in single cell eukaryotes like *Leismania*. In these disease-causing organisms, CRISPR-Cas induced DSBs are often deleterious because of the absence of NHEJ and the presence of numerous repeat regions in the genome that are used for SSA-mediated repair, resulting in deletion of the intervening sequence (Zhang and Matlashewski, 2019). The unique feature of SSA to promote recombination between tandem-repeated DNA sequences has found limited application for genome editing between inserted sequence repeats. For example, in human T-cells a SSA-mediated DSB repair mechanism resulted in the precise removal of a (previously inserted) reporter gene flanked by repeats without leaving any scar behind (Li *et al.*, 2018a). In some cases, it is unclear whether MMEJ or SSA has been used for repairing a DSB due to the intermediate size of the co-

delivered homology regions, *e.g.* in genome editing in zebrafish and human cells using ErCas12a (Wierson *et al.*, 2019).

5.1.2.2.4 Other repair pathways

Our understanding of damage repair processes is incomplete and the distinction between the above described pathways may in some cases be artificial. Intermediate or alternative repair pathways may be active under certain conditions, *e.g.* when other pathways are not available, as has been shown in studies on deletion mutants. Recent studies have also revealed that SSA becomes inefficient in mammalian cells when the distance between the DSBs and the repeats is increased to the 1–2 kb range and an alternative process, called break induced repair (BIR), acting over a much longer distance may take over the repair process (Hu *et al.*, 2019). Other mechanisms for DSB repair, including synthesis-dependent strand annealing (SDSA; Miura *et al.*, 2012) and single-stranded templated repair (SSTR) with oligonucleotides as donor (Richardson *et al.*, 2018), have been proposed in studies on various organisms or cells and under different conditions (see reviews in Rodgers and McVey, 2016; Danner *et al.*, 2017). In a number of clever studies using ssDNA oligonucleotides, SSTR was shown to involve the cellular components of the Fanconi anemia repair pathway, which is normally involved in interstrand cross-link repair (Richardson *et al.*, 2018). This pathway seems to direct repair away from the error-prone NHEJ pathway and towards template-mediated repair based on HR or SSTR. The precise action mechanisms of these repair pathways are not yet completely elucidated.

5.1.2.3 Strategies favouring repair through homologous recombination

Whereas NHEJ and several of the HDR pathways are known as error-prone repair processes, HR in principle results in the precise correction of DSBs in the genome. However, HR is limited to the S and G2 phases of the cell cycle and is not very efficient in most eukaryotic cells (except yeast). The inherent dominance of NHEJ for repairing DSBs means extra work and complications in genome editing experiments aimed at precise gene correction or gene addition (Danner *et al.*, 2017). Various approaches have, therefore, been attempted for increasing the frequency of HR over NHEJ (or other repair pathways).

Small molecules have been used to inhibit specific components of the NHEJ pathway in mammalian cells such as KU70/KU80, Lig4 or DNA-PKs (Danner *et al.*, 2017). Over-expression of Rad51 or of enhancers of Rad51 (Rees *et al.*, 2019b), or reducing the expression of key NHEJ proteins, such as 53BP1 (Canny *et al.*, 2018; Nambiar *et al.*, 2019), may also stimulate HR. Alternatively, synchronisation of the cell cycle before DSB induction and optimisation of the donor templates used may further increase the percentage of cells in a population that are repaired by HR (Danner *et al.*, 2017). Use of geminivirus replicons capable of enhanced donor template delivery to the cells has shown to increase HR tenfold in tomato using either TALEN or CRISPR-Cas (Čermák *et al.*, 2015). A CRISPR-Cas fusion with the *Agrobacterium tumefaciens* virD2 relaxase increased HR in rice by bringing the repair template in close proximity to the DSB (Ali *et al.*, 2020). Using a digital PCR assay that can distinguish HDR from NHEJ outcomes, Miyaoka *et al.* (2016) concluded that the HDR/NHEJ ratios were highly dependent on gene locus, nuclease platform, and (human) cell type. Despite these successes in mammalian cells and similar studies in plant cells, genome editing through HR in plants remains challenging (Vu *et al.*, 2019).

5.1.3 System dynamics of SDN-mediated genome editing

Of the SDNs available for inducing DSBs, CRISPR-Cas is the most versatile system due to its sole dependence on an easy to synthesise sgRNA for directing the Cas-sgRNA to the targeted site. The need to have a PAM site in the vicinity of the target site for the initial recognition steps can usually be addressed by use of Cas variants with different specificities. High-fidelity Cas variants have been developed and various other parameters should be optimised for successful genome editing in a particular organism without inducing other changes in the genome. In a broad experimental study with 7 Cas9 variants and 26,891 target sequences, the PAM specificity, activity and cleavage specificity of each Cas9 variant was assessed and used to build deep-learning computational models to predict its activity on any target sequence (Kim *et al.*, 2020). Studies on various aspects of the way the CRISPR-Cas system works have seen an exponential growth in the past 10 years and so did applications of this system for genome editing.

The generation of a DSB in a specific location of the genome through the activity of any of the SDNs is a trigger for activation of several DNA repair machineries. Some of these cellular repair systems may introduce mutations at the DSB and result in changes to the genomic sequence. Our understanding of the mechanism and dynamics of DNA repair systems has grown considerably over the past years but the complex interactions between repair factors remains incompletely understood. Whereas the targeted induction of one or more

DSBs in the genome can be achieved relatively easily in the cells of many organisms, the repair system by which the cell repairs the break is more difficult to control. This depends on many factors including cell type, cell cycle, chromatin accessibility and many others. Repair through the NHEJ mechanism is often the most prominent pathway, resulting in short deletions and sometimes short insertions at the break site. Two DSBs on the same chromosome may be repaired by removal of the intervening sequence. The efficiency of repair by homologous recombination with a donor template sequence without introducing other alterations is much lower, but has been successful in some cases.

In the past, NGTs relying on SDNs have been classified as SDN-1 (no donor template provided), SDN-2 (donor template with mismatches to the endogenous target locus, resulting in sequence replacement) and SDN-3 (donor template consisting of novel sequences, resulting in gene addition). Whereas SDN-1 was reported to rely on NHEJ and SDN-2 and SDN-3 on HDR (SAM, 2017), the increased understanding of these pathways as described above has shown that the situation is much more complex. Targeted gene insertion following a CRISPR-Cas or ZFN-induced DSB may also be achieved through NHEJ, using donor templates with or without homology arms to the target site, as shown in soybean (Bonawitz *et al.*, 2019). Also PITCh, using MMEJ (Nakade *et al.*, 2014), and HITI strategies, using NHEJ (Yamamoto and Gerbi, 2018), may be used for gene correction or insertion, as described earlier.

5.2 Site-specific recombinase-mediated engineering

Introduction: Site-specific recombination (SSR) systems operate in prokaryote and lower eukaryote hosts to mediate precise excision, inversion, or integration of defined DNA units for controlling a large variety of biological functions. SSR (also called recombineering in bacteria) is an *in vivo* method of genetic engineering for integrating exogenous ssDNA or dsDNA into the bacterial genome, catalysed by bacteriophage-encoded homologous recombination functions. SSR has also been employed in higher organisms, such as plants or animals, for the site-specific insertion of transgenes into hot spots or for the excision of reporter or selection genes. These recombinases require a target sequence or landing pad in the genome. SSRs introduce a DNA DSB without subsequently requiring the host DNA repair machinery to rearrange genomic DNA. Furthermore, recombinase-based approaches are also efficient in post-mitotic cells. Recent progress has extended the target-specificity and efficiency of SSR systems, opening up new applications.

Mechanism: SSR systems are composed of complex DNA-modifying enzymes, or combinations of enzymes, that bind, cleave, strand exchange, and rejoin DNA. Several distinct systems exist, mainly differentiated by their application field for genome editing, *i.e.* bacteria or eukaryotic organisms.

Recombineering in bacteria is based on recombination systems involving several independent enzymes that work in a stepwise process for the precise recombination of a donor sequence, flanked by short homology arms, into a target site.

In contrast, the application of SSRs in eukaryotes usually requires the introduction beforehand of the recombinase recognition sites into the desired position in the genome of a target organism. These recognition sites then subsequently function as the target sites for recombination with a donor template flanked by the same sites. One application of this SSR technique is the insertion of a transgene construct consisting of both a desired transgene and a selection marker or reporter gene flanked by recombinase recognition sites; the selection or reporter gene can then be removed from the genome in a subsequent recombinase-mediated cassette exchange (RMCE) step (Wirth *et al.*, 2007; Gidoni et al., 2008; Minorikawa and Nakayama, 2011; Bi *et al.*, 2018; Carpenter *et al.*, 2019; Wang *et al.*, 2019b). Another application is the stacking of transgenes in so-called safe genomic harbours[10] (Zhu *et al.*, 2017; Bi *et al.*, 2018; Collier *et al.*, 2018). Recently, through directed molecular evolution experiments designer recombinases were created that act as heterodimers and recognise target sites naturally occurring in (human) target genes (Meinke *et al.*, 2016; Lansing *et al.*, 2020). This obviates the need to first insert the recombinase recognition sequences at the target site.

Site-directed recombinase-mediated engineering: detailed mode of action

In bacteria like *E. coli*, recombinase-mediated engineering or recombineering is performed by introducing linear DNA substrates (single or double stranded DNA) containing the sequence to be inserted, flanked by ~50 bp homology arms to the target DNA, into cells expressing the phage-encoded recombination enzymes, *e.g.* the Red recombinase complex from bacteriophage lambda (Sharan *et al.*, 2009). If not already present in the host, the recombination enzymes can also be delivered as mobile recombination systems on a plasmid or replication-defective λ phage (Sharan *et al.*, 2009). The enzymes constituting the λ Red system, including an exonuclease, a ssDNA binding protein and a degradation-protective protein, recombine the linear DNA at the homologous target sequence in the chromosome or plasmid. This results in the perfect integration of the

donor DNA, *e.g.* an antibiotic resistance or other selection gene, thereby replacing the endogenous sequence and knocking down its expression. The same technique has been used for gene correction, deletion or insertion into the chromosomal genome or into plasmids (for vector construction) or bacterial artificial chromosomes (BACs) of several bacteria (*e.g.* Hu *et al.*, 2014). In other bacteria, similar, but often host-specific phage-based recombination systems have been developed (*e.g.* Corts et al., 2018; Wang *et al.*, 2019b).

Coupling the λ Red system with a CRISPR-Cas generated DSB increased the efficiency of bacterial recombineering and allowed the deletion and replacement of large chromosomal regions with a selectable marker gene (Jiang *et al.*, 2013; Pyne *et al.*, 2015).

The bacteriophage λ also harbours another well-studied recombinase system, acting site-specifically at attB recognition sites, the λ Integrase (λ int) (Landy, 2015). The λ int system is the founding member of the family of site-specific tyrosine recombinases, which are defined by a catalytic tyrosine residue used to attack and cleave the target DNA during a 'cut and paste' recombination process (reviewed in Meinke *et al.*, 2016). Due to their accuracy and relative simplicity, the site-specific tyrosine recombinase systems that have been the focus of most intensive studies in plants and animals are Cre/lox of bacteriophage P1 of E. coli, R/RS from the SR1 plasmid of the yeast *Zygosaccharomyces rouxii*, and FLP/FRT from the 2 *μ* plasmid of *Saccharomyces cerevisiae*. These systems function through the interaction of a recombinase homodimer (*e.g.* Cre, FLP, R) with a pair of identical recognition target sites. Each corresponding target site (34 bp lox and FRT; 31 bp RS) is palindromic, comprising 12-13 bp inverted repeats surrounding a 7-8 bp asymmetric spacer region that confers directionality to the site and, hence, to the recombination reaction (Gidoni *et al.*, 2008). For Cre, two homodimers assemble the recombination synapse in a homotetramer complex and recombine two loxP target sites in a concerted interaction mechanism without the need of any accessory factors (Meinke *et al.*, 2016).

Another family of recombinases, the serine recombinases of the resolvase/invertase type, such as the ØC31-integrase, provide a versatile alternative to tyrosine recombinases for genome engineering. In nature, these enzymes function as multi-domain protein complexes that coordinate recombination at attB sites in a highly modular manner. Different from tyrosine recombinases, the action of a serine recombinase dimer involves the creation of a DNA DSB, but, in contrast to SDNs, this occurs in the form of a concerted cleavage and re-ligation with the donor DNA present (Turan and Bode, 2011). Chimeric ZF or TALE recombinases have been engineered in which the native DNA binding domain of a serine recombinase has been replaced with a custom-designed ZF or TALE protein. Such TALER (TALE recombinases) were shown to recombine DNA into bacterial and mammalian cells (Gersbach *et al.*, 2011; Mercer *et al.*, 2012).

Inducible modifications: SSR can be used for site-specific gene insertion, deletion or replacement. It is a cut-and-paste precise DNA modification technique that does not create unwanted alterations at the target site. Recombineering with ssDNA oligonucleotides is limited to short alterations, whereas dsDNA donor templates may introduce larger sequences (Meinke *et al.*, 2016). In CRISPR-Cas recombineering, deletions up to 19.4 kb, encompassing several genes, and insertions up to 3 kb have been introduced in *E. coli* with varying recombination efficiency (1-47 %; Pyne *et al.*, 2015). In mouse embryonic cells, insertions of up to 120 kb from a human BAC clone have been realised using RMCE (Wirth *et al.*, 2007).

Except for some recombinase systems used in bacteria, such as λ Red, SSR often leaves behind a single recombinase recognition site at the recombination target site, such as a loxP or FRT sequence of 34 bp. Such heterologous recognition sites are not retained when using evolved designer recombinases (Lansing *et al.*, 2020), or using other techniques for their subsequent removal (Pyne *et al.*, 2015; Zolotarev *et al.*, 2019).

Result: SSR may result in gene correction by replacement of a defective sequence with a template with one or more basepair changes, gene knockout through sequence deletion or insertion of a disruption sequence, or gene knockin by insertion of one or more (cis-, intra- or trans-) genes at a specific target site (Meinke *et al.*, 2016). In some cases, recombinase recognition sites remain at the recombination site (as discussed earlier). The exact outcome is largely predictable as SSR systems usually function in a precise way.

Target organism(s): Recombineering using bacteriophage-derived recombinase systems based on homologous recombination was initially applied to *E. coli* (Sharan *et al.*, 2009; Pyne *et al.*, 2015) and later also to a number of other bacteria (*e.g.* Minorikawa and Nakayama, 2011; Santos and Yoshikuni, 2014; Aparicio *et al.*, 2020). SSR systems such as the Cre/loxP system were most frequently used in mammalian systems initially (mouse, human cells, *e.g.* Wirth *et al.*, 2007; Lansing *et al.*, 2020), whereas the FLP/FRT system had found widespread use in flies (Drosophila, *e.g.* Zolotarev *et al.*, 2019). However, in recent years both systems, as well as a number of other SSR systems, have found prevalent use in a large variety of eukaryotic organisms (Meinke *et al.*, 2016), including plants (*e.g.* Furuhata *et al.*, 2019). Cre/loxP recombination systems have also been

employed for large chromosomal rearrangements as part of the development of synthetic yeast strains (Shen *et al.*, 2018).

Known off-target effects: Strong expression of some recombinases may be toxic to cells, as shown for Cre. Similar Flp-mediated off-target effects in heterologous organisms have not been reported so far (Meinke *et al.*, 2016). The existence of pseudo-recognition sequences in the genome may lead to off-target recombination (Chalberg *et al.*, 2006). This has to be carefully assessed for every particular application in a specific organism, particularly for ensuring its safety in gene therapy applications in humans. Using designer recombinases on human cells, no off-target modifications were detected by PCR, despite the existence of potential off-target recombination sites (Lansing *et al.*, 2020).

Limitations and gaps in knowledge: Recombinase-based engineering is generally very precise, but suffers from a low efficiency, and therefore, often requires the introduction of a selectable marker for identification of the recombination success (Pyne *et al.*, 2015). Subsequent excision of the selectable marker sequence using SSR (except in bacterial recombineering) often leaves a scar present in the genome (the single copy of the recombinase recognition sequence), which may be undesirable. Expression of the recombinase may also need to be controlled, *e.g.* with the use of inducible promoters (Aparicio *et al.*, 2020). Although based on a somewhat older technology, this NGT still finds useful applications, *e.g.* in synthetic biology (Santos and Yoshikuni, 2014; Shen *et al.*, 2018; Aparicio *et al.*, 2020).

5.3 Site-specific DNA transposition

Introduction: Transposable elements or, in short, transposons are long-time known genetic elements that may constitute a significant part of a prokaryotic or eukaryotic genome, *e.g.* they occupy 45 % of the human genome and >80 % of the maize genome (reviewed in Muñoz-López and García-Pérez, 2010). These selfish sequences are capable of jumping themselves into other genomic sites by a cut-and-paste mechanism, often leading to gene disruptions or (partial) sequence duplications (Hedges and Deininger, 2007). They are powerful drivers of genetic change and have played a significant role in the evolution of many organisms' genomes (Feschotte and Pritham, 2007). Transposons have been used for transgene insertion, including for therapeutic uses, *e.g.* to treat age-related blindness in humans and for the engineering of CAR T cells to fight cancer (Hudecek *et al.*, 2017). However, as such insertions are semi-random, defined only by short recognition sites in the genome, the outcome is not always predictable.

For controlled uses in genome editing, transposases have been fused to various DNA binding factors to direct programmed transposon insertions to specific loci. This mimics natural systems in bacteria and archaea, where Tn7-like transposons have hijacked components of the type I CRISPR-Cas systems for RNA-guided transposition (Peters *et al.*, 2017). DNA transposition may be an alternative to HDR for the insertion of desired genes into specific locations without requirement for DSB-induced repair or a homology template.

Mechanism: DNA transposons can be targeted to a desired genomic site by fusing them to a site-directed DNA binding domain. DNA transposons essentially consist of a transposase gene that is flanked by two Terminal Inverted Repeats (TIRs) of 9 to 40 bp long. The transposase recognises these TIRs to perform the excision of the transposon DNA. The excised DNA is subsequently inserted into a new genomic location, thereby duplicating the target recognition site. Target site insertion occurs randomly at short recognition targets without the involvement of host factors, *e.g.* the popular Himar1 transposons derived from the horn fly *Haematobia irritans* insert themselves into TA dinucleotides. Interestingly, the transposase function may be separated from the TIR-flanked gene of interest for transgene insertion into host sites.

Site-specific DNA transposition: detailed mode of action

DNA transposons operate by making four nicks, two at the target site to generate a staggered break at the transposase recognition sequence and two at the TIRs flanking the transposon. At each end of the transposon, the 3' end of one strand of the transposon is then joined to the 5' extension of one strand at the target site, a ligation reaction catalysed by the transposase. Finally, the remaining short single strands are filled in and religated by host enzymes (non-replicative transposition). This repair process generates duplicated sequences directly flanking the ends of the integrated transposon, known as target site duplications (Dyda *et al.*, 2012).

Several studies have described fusing a transposase to a DNA binding domain to target transposition to specific sites. Fusing the Gal4 DNA binding protein to Mos1 (a Tc1/mariner family member) and *piggyBac* transposases increased the frequency of integration near Gal4 recognition sites (Maragathavally *et al.*, 2006). Fusion of DNA-binding zinc-finger or transcription activator-like (TAL) effector proteins to *piggyBac* enabled integration into specified genomic loci in human cells (Owens *et al.*, 2013). More recently, the C-terminus of

the hyperactive Himar1C9 variant was fused to the N-terminus of the (nuclease-deficient) dCas9 using a flexible protein linker. The so-called CasTn (Cas-transposon) system requires spontaneous dimerisation of the Himar1 domain with that of another fusion protein. The active Himar1 dimer is guided to a sgRNA-specific target locus by the dCas9 domains in the Himar-dCas9 dimer. The Himar1 transposon is then cut-and-pasted into the target gene at the nearest TA sequence distal to the 5' end of the sgRNA (Chen and Wang, 2019).

An alternative system is based on the cyanobacterial CAST (CRISPR-associated transposase) system which comprises a miniature type I or V CRISPR-Cas encoded within a Tn7-like transposon (Strecker *et al.*, 2019). The authors showed transposition of up to 10 kb donor sequences into a specific region of an *E. coli* plasmid several dozen bp from the PAM site, in some cases with very high (up to 80 %) efficiencies without selection. The donor sequences, flanked by the transposon left end (LE) and right end (RE), were provided on a donor plasmid distinct from the helper plasmid providing the transposition proteins. In some of these insertions also the donor plasmid had been unexpectedly integrated, which may be avoided with an optimised donor sequence (Strecker *et al.*, 2020).

Similarly, Klompe *et al.* (2019) employed a transposon from the bacterium *Vibrio cholerae* carrying a mini-type I CRISPR-Cas to integrate DNA downstream of a genomic target site complementary to a guide RNA. They showed the involvement of a cascade of distinct Cas gene products encoded by the transposon to accomplish replicative transposition as opposed to cut-and-paste transposition at a fixed position (~49 bp) downstream of the cascade complex binding. The authors proposed the INsert Transposable Elements by Guide RNA-Assisted TargEting (INTEGRATE) system as '*an opportunity for site-specific DNA integration that would obviate the need for double-strand breaks in the target DNA, homology arms in the donor DNA, and host DNA repair factors*'.

Inducible modifications: This type of NGT leads to the site-specific insertion of a transposon sequence into the genome at the transposon recognition sequence (*e.g.* TA for Himar1, TTAA for piggyback, etc.). When the transposase is provided on a separate plasmid or as mRNA, the transposon may be loaded with exogenous DNA sequences, *e.g.* a reporter or selection gene or any gene of interest, which would then be inserted into the genome. Donor cargo DNA of up to 10 kilobases, provided on a distinct plasmid vector, has been integrated efficiently into various genomic sites in *E. coli* (Strecker *et al.*, 2019). DNA transposition is not scarless as it also introduces the left and right transposon recognition sequences and duplicates the few insertion site basepairs (Strecker *et al.*, 2019).

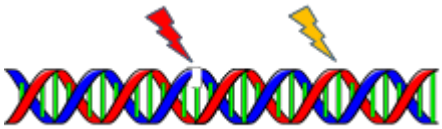
Result: Site-specific transposon insertion in a gene coding sequence results in gene knockout. Transposon integration may also be directed to the promotor region of a gene, which may lead to reduced (gene knockdown) expression of the protein encoded by the endogenous gene. Exogenous cis/intragenic or transgenic donor DNA may also be inserted site-specifically, flanked by the transposon recognition sites.

Target organism(s): As transposon movements do not require host factors, they can theoretically be used for site-specific integration into the genome of any prokaryotic or eukaryotic organism. So far, CasTn and Gal4-based transposition has been successfully used to insert transposons into *E. coli* plasmids (Maragathavally *et al.*, 2006; Chen and Wang, 2019). The INTEGRATE system was used for transposition into the *E. coli* circular chromosome (Klompe *et al.*, 2019). Transposition to mammalian cells using CasTn failed and may require further optimisation (Chen and Wang, 2019). ZF or TALE fusions were applied to human cells *in vitro* (Owens *et al.*, 2013). dCas9 fusions with several transposons, including *piggyBack* and *Sleeping Beauty*, demonstrated targeted transgene insertion into human cells, however, with limited specificity (Hew *et al.*, 2019; Kovač *et al.*, 2020).

Known off-target effects: Background activity of the transposase, even when fused to a DNA binding domain, has been reported in *E. coli* and in human cells and may result in integration of the transposon in unwanted transposon-recognition sites (Strecker *et al.*, 2019; Chen and Wang, 2019; Hew *et al.*, 2019; Kovač *et al.*, 2020). The CAST system was reported to generate an on-target accuracy >95 % across dozens of target sites (Strecker *et al.*, 2019), and also the INTEGRATE system produced few off-target transpositions in *E. coli* (Klompe *et al.*, 2019).

Limitations and gaps in knowledge: DNA transposition to a target site in the genome results in integration of the transposon sequence harbouring any exogenous sequences inserted between the transposon TIRs. The efficiency of the promising CasTn system has only been shown experimentally using bacterial reporter plasmids (Chen and Wang, 2019). Further optimisation of this non-viral gene transfer technique and monitoring of the precision of the outcome will be required for practical applications in higher organisms, particularly for therapeutic applications (Kumar, 2020).

6 Group 2: Genome editing without DNA double-strand break



For targeted genome editing the generation of a DSB is not an absolute prerequisite. Before the discovery of the SDNs, small DNA sequences called oligonucleotides have been used successfully to make short changes at a predefined location in the genome. This oligo-directed mutagenesis (ODM) technique was later shown to be more efficient after the induction of a DSB by an SDN (Rivera-Torres and Kmiec, 2016). This technique is described below first (Section 6.1).

Given the potential undesirable and uncontrolled consequences of introducing DSBs in living cells, which induce a battery of DNA repair pathways that are difficult to control, genome editing following single-strand breaks (SSBs or nicks in the DNA) has been proposed as a less mutagenic alternative to DSBs. All classes of SDNs can be converted into nickases through mutagenesis of their catalytic sites, which has resulted in enhanced HR and fewer off-target effects (Davis and Maizels, 2011; Metzger *et al.*, 2011; Ramirez *et al.*, 2012; Wu *et al.*, 2014; Fu *et al.*, 2019). In these experiments, either single nicks in the DNA or double nicks have been generated to stimulate editing (Ran *et al.*, 2013). Individual nicks are predominantly repaired by the high-fidelity base excision repair pathway (BER) (Roldán-Arjona *et al.*, 2019). Although double nicks are usually designed to opposite DNA strands, paired double nicking of the same DNA strand showed effective in gene knockin from a co-delivered donor template in human cells (Hyodo *et al.*, 2020).

The use of catalytically impaired SDN variants (nickases) for genome editing will not be described as a separate NGT in this study, as their use is similar to the original versions from which they are derived. The only difference is that they induce a ssDNA break instead of a DSB, which is less detrimental to the cell (for references: see previous paragraph). More important is that these SDN variants, and particularly the nCas (nickase Cas) and dCas (dead Cas, with both nuclease domains inactivated) variants, can be harnessed for various other functions by fusion to distinct effector proteins. An indefinite list of novel opportunities for genome editing has seen the light in recent years and these are described in Sections 6.2 and 6.3.

6.1 Oligonucleotide-directed mutagenesis (ODM)

Introduction: ODM, also called targeted gene repair, is a genome editing technique that long preceded the discovery of the techniques utilising site-directed nucleases. It was first used and optimised in bacteria and yeast models in the mid-eighties, before being applied to animals and plants (reviewed in Rivera-Torres and Kmiec, 2016). Using oligonucleotides carrying one or more mismatches compared to the target region as donor template, the technique makes short basepair changes at specific target sites in the genome with low efficiency.

Mechanism: The technique uses short single-stranded oligonucleotides, chimeric DNA/RNA oligonucleotides (called chimeraplasts and forming a hairpin structure), or even double-stranded DNA molecules, homologous to a region of interest but centrally containing the desired mutation(s). The oligonucleotides are chemically synthesised and delivered to cells using electroporation, PEG-mediated techniques, microinjection or biolistic bombardment. The mechanism was initially thought to involve annealing of the oligonucleotide with the genomic DNA during replication, followed by incorporation of the mutation via the mismatch repair (MMR) pathway (Komor *et al.*, 2018). The MMR pathway is another DNA damage repair pathway distinct from DSB-induced DNA repair, which mainly deals with replication errors such as base:base mismatches or insertion/deletion loops (Li *et al.*, 2016c). However, inhibition of components of the MMR pathway enhanced gene editing efficiency with oligonucleotides in animal cells. It has therefore been hypothesised that the ODM mechanism may involve the creation of a DSB and repair by a HDR pathway (Papaioannou *et al.*, 2012).

Oligonucleotide-directed mutagenesis: detailed mode of action

Single-stranded DNA oligonucleotides have been succesfull in inducing short alterations in the DNA, although at low frequency. To increase the mutation frequency, DNA/RNA chimeroplasts proved useful. Within such chimeraplasts, the DNA sequence binds to its genomic target while the RNA part improves hybridisation of the DNA sequence. The resulting genomic DNA:chimeric oligonucleotide heteroduplex is recognised by the endogenous cellular mismatch repair machinery and repaired, albeit infrequently, using the chimeraplast as a

template instead of the genomic DNA (Cole-Strauss et al., 1996; Beetham *et al.*, 1999). Later, the use of modified nucleotides or other end-protective chemistries have further improved ODM's efficiency (Sauer *et al.*, 2016). Much of the knowledge on the mode of action of ODM has been based on studies with bacteria and yeast (Huen *et al.*, 2006). The involvement of mismatch repair activity in the creation of the mutation in animals or plants is not completely clear (Dekker *et al.*, 2006; Wang *et al.*, 2006; Komor *et al.*, 2018). In the alternative 'replication model' of gene repair the oligonucleotide becomes incorporated into the newly synthesised strand of DNA during formation of a D-loop at a replication fork, which is then repaired through a second round of replication (Rivera-Torres and Kmiec, 2016).

The use of chromatin-relaxing chemicals, inhibiting histone deacetylation, enhances the frequency of gene repair with oligonucleotides in plants (Tiricz *et al.*, 2018), as well as the induction of DSBs in the neighbourhood of the target site by site-directed nucleases like HE or ZFN (Papaioannou *et al.*, 2012; Sauer *et al.*, 2016; Komor *et al.*, 2017). In conclusion, gene editing of single-base mutations through ODM initiates with the annealing of the oligonucleotide to the target sequence through complimentary base pairing (step 1). Hybridisation of the oligonucleotide (with a central mutation) creates a mismatched basepair(s) which acts as a triggering signal for the cell's repair enzymes. It seems that the subsequent pathway may be distinct (or optional) in different cells and organisms, involving direct mismatch repair or strand invasion in the presence or absence of DNA replication activity (Parekh-Olmedo *et al.*, 2005; Papaioannou *et al.*, 2012; Rivera-Torres and Kmiec, 2016). The process involved has been considered a type of gene repair and not homologous recombination (Breyer *et al.*, 2009), although components of the HR pathway participate (Papaioannou *et al.*, 2012). Richardson *et al.* (2018) reported that single-strand DNA (ssDNA) templates are incorporated through a pathway that is RAD51-independent but requires the Fanconi anemia pathway, although the exact mechanism is still to be elucidated.

A different class of oligonucleotides used for genome editing are the triplex-forming oligonucleotides (TFOs) and peptide-nucleic acids (PNAs), which are able to form DNA-DNA-DNA homotriplex structures or PNA-DNA-PNA heterotriplexes with genomic DNA, respectively (recently reviewed in Economos *et al.*, 2020). Such unusual structures are repaired, at least partly, by the nucleotide excision repair (NER) pathway, which is a non-mutagenic pathway for single-strand break repair in mammalian cells, distinct from those used for DSB repair or MMR. In the presence of a ssDNA donor molecule with a mismatch, co-administered to the cells, a recombination event may occur, incorporating a permanent sequence modification in the genome. Successful *in vivo* application of PNAs to treat various diseases in mouse models has provided insights in the potential for curative therapeutic application of PNAs to monogenic disorders (Economos *et al.*, 2020).

Inducible modifications: The alterations induced by ODM are limited to single nucleotide substitution, or insertion or deletion of one or a few nucleotides according to the design of the donor oligonucleotide. The oligonucleotide may be designed to contain one or a few mismatches in a central position amidst an otherwise homologous sequence compared to the target sequence, or it may have a nucleotide less or more. Insertions up to 4 bp have been obtained with low efficiency in mouse embryonic cells (Dekker *et al.*, 2006).

Result: Site-specific alteration of one or a few nucleotides in a gene sequence, which may result in gene correction, *e.g.* correction of a genetic disease or creation of a more (or less) powerful protein. An alternative outcome is disruption of gene function, *e.g.* by creation of a stop codon as a result of basepair replacement or of a frameshift due to deletion or insertion of a basepair, which may result in a non-functional protein.

Target organisms: ODM has been applied to plants (*e.g.* in maize, tobacco, rice, wheat and rapeseed), animals (mammals), fungi (yeast) and bacteria (Rivera-Torres and Kmiec, 2016).

Known off-target effects: The presence of off-target mutations has not been well studied for this NGT (Sauer *et al.*, 2016). It likely depends on the occurrence of alternative homologous sequences between the oligonucleotide donor and the targeted genome. Careful selection of the oligonucleotide used may prevent off-target effects.

Limitations and gaps in knowledge: ODM is suffering from a low and variable efficiency. Such unpredictable outcomes have made ODM (in its original form) unsuitable for most therapeutic uses. In combination with TFOs or PNAs, it may at some point find its way into human gene therapies. ODM has been particularly optimised for use in plant genome editing, with some successes in maize, tobacco, rice, wheat and rapeseed, despite requiring extensive single-nucleotide polymorphism screening to discover the desired genetic changes (Sauer *et al.*, 2016). In combination with site-directed nucleases inducing DSBs, the efficiency of ODM is greatly increased and may find applications in plant or animal mutagenesis as well as in human gene therapy (Economos *et al.*, 2020).

6.2 Base editing

Introduction: Base editors were developed to overcome the limitations of the CRISPR/Cas9 system for introducing point mutations in the genome: creating DSBs in the genome that are repaired by homology directed repair (HDR) or non-homologous end-joining (NHEJ) mechanisms can be not very efficient and/or introduce indels and chromosomal rearrangements (Ran *et al.*, 2017). Cytosine base editors were first developed in 2016 (Komor *et al.*, 2016; Nishida *et al.*, 2016); adenine base editors followed in 2017 (Gaudelli *et al.*, 2017). Base editing introduces single nucleotide changes from C to T or from A to G, with a few exceptions described below, at specific target sites in the genome. Base editors are constantly improved and new applications are continuously being developed. Very recently base editors capable of simultaneously deaminating cytosines and adenines at the same target site were developed (Grunewald *et al.*, 2020; Zhang *et al.*, 2020; Li *et al.*, 2020).

Mechanism: A base editor is a fusion of a catalytically inactive or modified Cas domain and a cytosine or an adenosine deaminase domain, which is then guided to the target site by CRISPR guide RNA (sgRNA) (Gaudelli *et al.*, 2017; Komor *et al.*, 2016). The use of a catalytically inactive or modified Cas domain (to have nickase activity in the strand opposite to the target base, with different modifications to improve or decrease PAM site specificity) involves no break or a single-strand break in the target DNA. The binding of the Cas9-sgRNA complex to the target DNA creates an 'R-loop' where a stretch of DNA bases is unpaired. The cytosine or adenine deaminase domain (or both, in the case of dual cytosine and adenine base editors) acts on this 5–8 nucleotides long stretch of single-strand DNA. The use of a nickase Cas has the purpose of directing cellular DNA repair mechanisms to correct the non-edited strand and therefore increase base editing efficiency (Komor *et al.*, 2016). The eukaryotic mismatch repair mechanism (MMR), which normally acts on newly synthesised DNA, uses the nicks to direct removal and re-synthesis of the non-edited strand (Komor *et al.*, 2016; Heller and Marians, 2006; Pluciennik *et al.*, 2010).

Base editing: detailed mode of action

There are two types of DNA base editors: cytosine base editors (CBEs) which can be used to modify C into T and adenine base editors (ABEs) which can be used for A to G modification (Hua *et al.*, 2018; Liu *et al.*, 2018a; Qin *et al.*, 2020; Zong *et al.*, 2017). Three versions of a base editor capable of simultaneous C to T and A to G modifications using the same guide RNA were also created recently (Grunewald *et al.*, 2020; Zhang *et al.*, 2020c). Different generations and versions of CBEs have been developed using different Cas9 variants, deaminases, linkers and domains with additional functions, with the purpose of increasing base editing efficiency and reducing off-target effects due to both Cas9 domain selectivity and off-target base deamination (Tan *et al.*, 2019). The conversion from U (operated by the deaminase domain) into T was identified as a source of incorrect base modification and indels formation and was addressed by the addition of one or two uracil DNA glycosylase inhibitor (UGI) domains in different positions (Komor *et al.*, 2016). Gam, a bacteriophage Mu protein that binds DSBs also greatly reduces indel formation and improves product purity (Komor *et al.*, 2017). Other aptamers were also used to facilitate simultaneous base conversions (C-T and A-G), *e.g.* MS2, PP7, COM and boxB (Ma *et al.*, 2016; Zalatan *et al.*, 2015). ABEs have also been developed using different Cas9 and deaminase variants, but exhibited lower off-target editing than CBEs and were therefore modified less extensively. For both ABEs and CBEs codon optimisation and nuclear localisation signals are crucial for improving their efficiency and precision (Koblan *et al.*, 2018).

ABEs and CBEs can be used for multiplex gene editing using Cas9 and Cas12a (Cpf1) orthologs (Kim *et al.*, 2017b; Hirano *et al.*, 2016; Webber *et al.*, 2019; Zetsche *et al.*, 2017). Recently, a new type of bacterial deaminase (called double-stranded DNA deaminase toxin A or DddA) has been characterised that acts on double-strand DNA (Mok *et al.*, 2020). The deaminase domain was split into two halves for reduced toxicity, and fused to a TALEN pair for highly specific C•G to T•A conversions in human nuclear, but also mitochondrial DNA (Mok *et al.*, 2020). As genome editing in mitochondrial DNA was previously not possible due to challenges in delivering a sgRNA into mitochondria, this (non-CRISPR-based) tool may have future potential for the treatment of human mitochondrial disorders.

TAM and CRISPR-X platforms have been created specifically for creating libraries of point mutations localised to targeted regions of mammalian genomes. In both systems, dCas9 is fused to a cytidine deaminase domain that converts C into A, G, or T and G into A, C, or T (Ma *et al.*, 2016; Hess *et al.*, 2016). The TAM system exhibits strong activity (>20 %) with transitions and transversions from cytidine and guanine to the other three bases (Ma *et al.*, 2016). CRISPR-X exhibited a window of catalytic activities between −50 and +50 basepairs from the PAM sequence (Hess *et al.*, 2016).

Inducible modifications: Single base C to T or A to G conversion in DNA sequences. TAM and CRISPR-X platforms convert C into A, G, or T and G into A, C, or T (Hess *et al.*, 2016; Ma *et al.*, 2016). Dual base editors can make C to T and A to G base changes at the same target site simultaneously, considerably expanding the editing capacity. With the dsDNA-specific DddA deaminase C•G to T•A conversions are possible (Mok *et al.*, 2020).

Result: Site-specific alteration of single nucleotides in a gene sequence, which may result in gene correction, *e.g.* correction of a genetic disease or creation of a more (or less) powerful protein (for examples of application in crops see Mishra *et al.*, 2020). An alternative outcome is disruption of gene function, *e.g.* by creation of a stop codon as a result of a basepair replacement. As an example, CRISPR-STOP and iSTOP platforms use CBEs for introducing stop codons in gene coding sequences, offering an alternative method to knockout genes without relying on the unpredictable mutations resulting from introduction of DSBs (Kuscu *et al.*, 2017; Billon *et al.*, 2017). Base editing may be used for inducing exon skipping: the CRISPR-SKIP platform utilises CBEs for programming exon skipping by mutating target DNA bases within splice acceptor sites. To facilitate the identification of exons that can be skipped with the various base editors, the group developed a web-based software tool (Gapinske *et al.*, 2018). Another possibility is its use for creating libraries of point mutations and for directed evolution (TAM and CRISPR-X platforms: Ma *et al.*, 2016; Hess *et al.*, 2016; STEME dual cytosine and adenine base editor platform: Li *et al.*, 2020). Molecular recording systems, like for example lineage tracing of the cells from an embryo to a mature organism (McKenna *et al.*, 2016; Chan *et al.*, 2019), are listed as possible applications for dual cytosine and adenine base editors (Grünewald *et al.*, 2020).

Target organisms: Bacteria (Wang *et al.*, 2020b), viruses, plants (applied *e.g.* in maize, wheat, rice, tomato, watermelon, potato, cotton: Li *et al.*, 2018b; Lu and Zhu, 2017; Tang *et al.*, 2019; Zong *et al.*, 2017; Tian *et al.*, 2018; Veillet *et al.*, 2019; Qin *et al.*, 2020), fungi (Bae *et al.*, 2020), animals and animal cells (zebrafish, mouse, rat, rabbit and pig: Kim *et al.*, 2017c; Ma *et al.*, 2018; Liu *et al.*, 2018b; Xie *et al.*, 2019), human cells and human embryos (Kim *et al.*, 2017b; Liang *et al.*, 2017b; Grünewald *et al.*, 2020; Zhang *et al.*, 2020c).

Known off-target effects: Off-target mutation can be sgRNA-dependent or independent. The presence of off-target mutations has been studied for some base editor versions in mouse and human cells and in rice (Zuo *et al.*, 2019; Jin *et al.*, 2019). It largely depends on the base editor version used and on the delivery method. DNA-free lipid- mediated delivery of ribonucleoproteins displayed lower off-target effects. Adenine base editors exhibited lower off-target editing than cytidine base editors. For cytidine base editors, to minimise the off-target mutations it is necessary to optimise the cytidine deaminase domain (codon usage, expression level and duration) and/or UGI components (Zuo *et al.*, 2019; Jin *et al.*, 2019). A careful selection of the most appropriate base editor version and progress in the field are expected to limit off-target effects.

In addition to off-target effects on genomic DNA, sgRNA-independent editing of RNA transcripts was observed in cells expressing high levels of cytidine base editors, but not in cells expressing low levels of the editor protein (Rees *et al.*, 2019a; Zhou *et al.*, 2019a; Grünewald *et al.*, 2019). However novel variants of cytosine base editors have been created to address this issue (Zhou *et al.*, 2019a; Grünewald *et al.*, 2019; Wang *et al.*, 2018b).

Dual cytosine and adenine base editors have similar or lower off-target editing activity on both DNA and RNA (Grünewald *et al.*, 2020; Zhang *et al.*, 2020c).

Limitations and gaps in knowledge: There is a dependence on the presence of a PAM site within an editing window from the base that needs to be edited (NGG PAM for SpCas9). To overcome this limitation several research groups have developed ABE and CBE base editors using Cas9 variants which recognise other PAMs and in some cases almost entirely removed dependence on a requisite PAM (Endo *et al.*, 2019; Hua *et al.*, 2018; Nishimasu *et al.*, 2018; Qin *et al.*, 2020; Wang *et al.*, 2019c; Walton *et al.*, 2020). Some Cas variants increased the editing efficiency (Kim *et al.*, 2017b). In other cases, PAM relaxation comes at the expense of a reduced relative efficiency and can also reduce specificity (Walton *et al.*, 2020; Nishimasu *et al.*, 2018; Kleinstiver *et al.*, 2019). A Cas12a-based cytidine deaminase base editor (Li *et al.*, 2018c) favours T-rich PAMs. Besides using different Cas domains, this limitation can be addressed by using base editors that have a different structure or cleavage mechanism, resulting in a different target window distance from the PAM site (Adli, 2018; Gehrke *et al.*, 2018).

Editing window: cytosine deaminase base editors can potentially edit any C that is present in the activity window that is either narrow or wide (4–5 nucleotides, up to 9), at asymmetric frequency distributions, depending on the type of deaminase used (Komor *et al.*, 2016; Zong *et al.*, 2018, Huang *et al.*, 2019b). Efforts were made to generate high-precision base editors with narrow catalytic windows that can precisely edit a single cytidine residue within the catalytic window with high accuracy and efficiency (Tan *et al.*, 2019). Dual

cytosine and adenine base editors have editing windows similar to the single cytidine or adenine editor domains they harbour, perhaps a few nucleotides wider or narrower, depending on the editor (Grünewald *et al.*, 2020; Zhang *et al.*, 2020c).

So far base editors can only effect transition mutations (C → T, A → G), not transversion mutations (C to A or G and A to C or T), with the exception of the TAM and CRISPR-X platforms described above and the dsDNA-specific Ddda deaminase. However, very recently new base editing platforms capable to convert C to G in mammalian cells and C to A in *E. coli* have been developed (Kurt *et al.*, 2020; Zhao *et al.*, 2020). Further developments and optimisation of these platforms is needed to increase their efficiency and understand their more complex mechanism of action.

There is also evidence that DNA methylation inhibits base editing efficiency, although some base editor versions have higher editing efficiency in highly methylated regions (Wang *et al.*, 2018b).

6.3 Prime editing

Introduction: Genome editing following DSB formation has some limitations, including the dominant repair through imperfect end joining and the risk for off-target modifications. To address these shortcomings, prime editing was developed, which allows to engineer all 12 base substitutions, as well as short insertions, deletions, and their combinations, without requiring DSBs or the provision of homology templates. Instead, it uses a partially impaired Cas protein producing ssDNA nicks fused with a reverse transcriptase (RT) that converts the extended sgRNA template into the desired DNA sequence, which is then incorporated into the genome. Prime editing is a 'search-and-replace' genome editing tool that is able to make modifications at locations further away from the PAM site than any other CRISPR-Cas based technique, which increases its potential to make any desired modification in the genome.

Mechanism: Prime editors (PEs) use a reverse transcriptase (RT) fused to an RNA-programmable nickase and a prime editing extended guide RNA (pegRNA). The pegRNA recognises the target sequence, and the nickase induces a ssDNA break. The nicked DNA strand subsequently functions to prime the reverse transcriptase reaction for copying the RNA sequence of the sgRNA (with its mismatch(es)) into DNA. The DNA is repaired by endogenous repair systems, substituting the original sequence with the new pegRNA-derived sequence.

Prime editing: detailed mode of action

Anzalone *et al.* (2019) fused the Cas9 H840A nickase (Cas9 with the single amino acid mutation H840A in the HNH domain) to the Moloney murine leukemia virus (M-MLV) RT to generate a ssDNA break at a target locus identified by the pegRNA. The pegRNA is designed to provide both the 'search' part directing Cas9 to the target site and an RT template of ~10-30 nucleotides for directly 'replacing' the genetic information of the target genomic locus with the reverse transcribed sequence of the RT template. The reverse transcriptase reaction is primed by the nicked (non-template) DNA strand containing a free 3'-hydroxyl group which hybridises to a complementary primer binding site of ~13 bp on the pegRNA. The RT then produces DNA complementary to the sequence in the 'replace' part of the RNA guide, resulting in a branched DNA intermediate. Endogenous DNA repair mechanisms can then remove the 5'-flap that contains the unedited sequence, thereby fixing the edit in place. The use of a second RNA guide to direct a cut to the original intact strand may increase the chances of repair of this non-edited strand according to the newly generated reverse transcribed sequence. Anzalone *et al.* (2019) systematically analysed and optimised the different components of this technique, including use of an engineered RT, optimisation of the length of the RT-template and primer binding site of the pegRNA, and use of a second sgRNA with spacers that match the edited strand but not the non-edited one, thereby favouring nicking the second strand only after editing the first strand.

A distinct but related approach has been used by Li *et al.* in rice (2019b), where the extended sgRNA provides the RNA template for direct HR into the rice genome without requiring its conversion into DNA by a Cas-RT fusion protein.

Inducible modifications: In the only paper describing this NGT so far, Anzalone *et al.* (2019) demonstrated that all 12 possible types of single nucleotide alterations (4 bases can each be changed to the three other bases) can be introduced to target genes at locations ranging from 3 bp upstream to 29 bp downstream of a PAM. It can also induce changes in the PAM sequence itself. Furthermore, prime editing can perform insertions at least up to 44 bp and deletions up to 80 bp in human cells and mouse cortical neurons.

Result: Prime editing can result in the site-specific alteration of one or a few nucleotides (< 100 bp) in a gene sequence, which may result in gene correction, *e.g.* correction of a genetic disease or creation of a more (or less) powerful protein. An alternative outcome is disruption of gene function, *e.g.* by creation of a stop codon

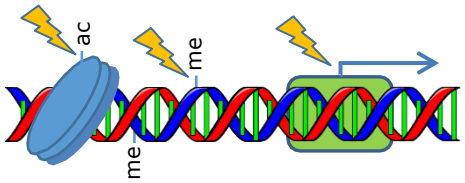
as a result of basepair replacement or of a frameshift due to deletion or insertion of one or more basepairs, which may result in a non-functional protein. Large deletions or insertions (*e.g.* >100 bp) may not be possible (Jinek, 2019).

Target organism(s): Prime editing has only been tested on human cells and mouse cortical neuronal cells, and partially in yeast (Anzalone *et al.*, 2019). There is, however, no a priori reason why it would not work in other organisms. The distinct approach proposed by Li *et al.* (2019b) was tested in rice and could work in other plants.

Known off-target effects: Prime editing is reported to offer much lower off-target activity than Cas9 at known Cas9 off-target loci, far fewer byproducts and higher or similar efficiency compared to Cas9-initiated HDR (Anzalone *et al.*, 2019). In contrast to base editors, prime editing is more efficient at modifying only a single base when other similar bases (*e.g.* a C) are nearby, or when there is no PAM sequence close to the target site.

Limitations and gaps in knowledge: Despite the promising potential of prime editing at producing precise edits without unwanted indel byproduct, which is particularly important for human applications, the technique has been applied to a limited number of cells in one laboratory only. Future work will have to confirm the results obtained and a significant amount of orchestration of the action of the different components of the system may be needed for predictable results. The nCas-RT fusion is also quite large in comparison to Cas9 and delivery to certain cells may therefore be challenging (Platt, 2019).

7 Group 3: Editing of the epigenome



Editing the epigenome[11] is another approach for altering the functional activity of the genome (Kungulowski and Jeltsch, 2016). Epigenome is the term used to indicate the overall chromatin[12] composition, which marks the genome of any given cell by adding covalently linked chemical compounds on DNA (such as a methyl group on cytosine nucleotides) or on histone proteins (such as acetyl or methyl groups on lysine residues of core histone tails).

In a broad sense, epigenome editing can be achieved by directly modifying the epigenetic marks on DNA or histones or indirectly by affecting the transcriptional regulation through activators and repressors (Kungulowski and Jeltsch, 2016; Porter *et al.*, 2019; Miglani *et al.*, 2020). Changing the epigenome does not involve breaking the target DNA molecule as it happens in reactions catalysed by nucleases, transposases (Stepper *et al.*, 2017) or recombinases (Nomura, 2018).

Site-directed epigenome editing systems are essentially built by linking a DNA recognition domain (from ZFN, TALEN or CRISPR-Cas) with the active domain of a chromatin-modifying enzyme, thereby delivering the epi-effector enzyme to the intended locus (Kungulowski and Jeltsch, 2016). The epi-effectors can be divided into two categories: the chromatin-modifying enzyme can act directly on the DNA or histones to modify their epigenetic state (Section 7.1) or it can indirectly change the epigenetic state through the action of transcriptional modulators like activators or repressors (Section 7.2).

7.1 Site-specific modulation of the epigenetic state

Introduction: The epigenetic state of DNA sequences affects gene activation or repression. The expression 'synthetic epigenetics' is sometimes used to refer to the precise addition or removal of chromatin marks through delivery of sequence-specific molecular tools that enable a change of the epigenetic state of selected regions (Stepper *et al.*, 2017). The targets for these precise alterations can be either DNA or histones, but the outcomes of these alterations are similar.

Chromatin remodelling occurs through changes in epigenetic marks. This signals to the transcriptional machinery which genomic regions are made accessible in a cell-type and developmental-specific manner under certain environmental stressors (Miglani *et al.*, 2020). As a result of chromatin remodelling the DNA-protein complex can be 'opened' so that specific genes are expressed. After cell differentiation, the epigenetic identity is maintained in next cell generations by the cross-talk between multiple and likely redundant epigenetic marks, of which the specific roles are not yet completely understood. Also the transcriptional activity of specific gene sequences can strengthen the preservation of the epigenetic state (Jones, 2012; Bergman and Cedar, 2013; Carey, 2012; Jurkowski *et al.*, 2015).

Mechanism: With regard to their mode of action, DNA methylation and histone modifiers are described separately.

DNA methylation modifiers

DNA methylation consists of the covalent addition of a methyl group to the C5 position of mostly cytosine residues in DNA by a DNA methyltransferase. In mammals, this occurs usually at CpG dinucleotides, of which around 75 % are methylated. However, clusters of CpG repeats found at promoter regions and first exons of

[11] The epigenome is the term used to indicate the overall chromatin composition, which marks the genome of any given cell by adding covalently linked chemical compounds on DNA or on histone proteins, thereby potentially affecting the transcription of the DNA into RNA. Whereas the term genome essentially refers to the DNA sequence of a cell, the epigenome refers to the additional reversible modifications of the DNA sequence which allow the same sequence to express different functionalities, *e.g.* in line with cell differentiation or in reaction to environmental stimuli.

[12] In eukaryotes, DNA is tightly wound into a complex called chromatin, in which DNA is packed around proteins called histones to form nucleosomes. Nucleosomes are then grouped into nucleosome clusters, which in turn are grouped to form chromatin fiber, which is further wrapped into higher structures, ultimately forming chromosomes.

genes are frequently unmethylated (Tost, 2010). In plants the cytosine can be methylated at CpG, CpHpG, and CpHpH sites (H indicates adenine, cytosine or thymine). In rice, *A. thaliana* and other plant species, like in mammals, most of the cytosines in the CpG dinucleotides frequently associated with the promoter region or the whole gene sequence of actively transcribed genes are unmethylated (Ashikawa *et al.*, 2001). Extensive methylation of gene promoter regions is associated with condensed chromatin structure and transcriptional repression (Hilton and Gersbach, 2015). DNA methyltransferases (DNMT) use S-adenosyl-L-methionine (SAM) as methyl donor, leading to the formation of 5-methylcytosine (5mC).

For the site-specific modification of the methylation marks on DNA, the catalytic domain of a DNA methyltransferase is fused to a site-specific DNA binding protein. Alternatively, the DNA binding protein may be fused to a demethylase, resulting in the site-specific removal of methyl groups from the targeted DNA.

DNA methylation modifiers: detailed mode of action

ZFs and TALEs have shown potential to direct DNA methylation in a targeted, locus-specific manner. *E.g.* the C-terminal parts of Dnmt3a and Dnmt3L were fused through a 15-30 amino acids linker and then linked to an artificial zinc finger (Cys2-His2) (Siddique *et al.*, 2013) or TALE (Bernstein *et al.*, 2015) for *de novo* methylation at specific loci in human cells. Through such approaches, cytosines at other sites in the genome may become methylated as well. *E.g.* DNA methylation by TALE-Dnmt decreased with distance from the transcription start site of the targeted locus, while methylation at CpG islands of a reference gene on another chromosome remained unaffected (Bernstein *et al.*, 2015). DNA methylation of CpG islands initiated in proximity of a transcriptional start site within targeted promoters can spread up to 1200 bp (Stepper *et al.*, 2017). Interestingly, an increase in methylation was observed around 25 bp upstream and 40 bp downstream of the PAM site, while 20–30 bp of the dCas9 binding site itself were protected against methylation, possibly by dCas9 occupancy.

In plants, the *fwa* epiallele in *Arabidopsis* displays a loss of 5mC (methylated cytosine) at the *fwa* promoter leading to delayed flowering. The ZF-targeted methylation effector SUVH9 re-introduced promoter methylation at the *fwa* epiallele, thereby silencing the *fwa* gene and restoring the heritable early flowering phenotype (Papikian *et al.*, 2019).

Several approaches have been used to enhance the specificity of these DNA methylation modifiers for single CpG sites, avoiding off-target methylation, *e.g.* through *in vitro* directed evolution selection (Chaikind and Ostermeier, 2014; Nomura and Barbas, 2007).

To achieve target demethylation the catalytic domain of human TET1 (or TET protein family members) has been used. Ten-Eleven Translocation (TET) family is a group of demethylases catalysing the oxidation of 5-methylcytosine (5-mC) to 5-hydroxymethylcytosine (5-hmC), then stepwise into further oxidation products, which finally are converted to cytosine by one of two endogenous processes (Lu *et al.*, 2015; Wu and Zhang, 2017). A ZFN or plant-optimised CRISPR/dCas9-based SunTag[13]–TET1cd system fused to the human demethylase TET1 domain could induce demethylation at the *fwa* gene, overexpression and the late flowering phenotype similar to the *fwa* epiallele (Gallego-Bartholomé *et al.*, 2018). By direct fusion of the TET1 catalytic domain to engineered TALE repeats it was possible to achieve demethylation of specific CpG positions in promoter regions and thereby increase the expression of endogenous human genes (Maeder *et al.*, 2013a). A similar approach was used by fusing TET-1 and TET-2 with zinc finger domains (Chen *et al.*, 2014; Gallego-Bartolomé *et al.*, 2018).

The DNA recognition and binding characteristics of a catalytically inactive (dead) sgRNA-dCas9 system have also been exploited to deliver epigenetic changes to specific DNA targets. Fusion of Dnmt3a (*de novo* methylation) or TET1 (demethylation) with dCas9 enabled to target and edit the DNA methylation at promoter sequences in mice and affect respectively gene silencing or activation of an endogenous reporter *in vitro* and *in vivo* (Liu *et al.*, 2016). Employing the catalytic domains of two DNA methyltransferases (Dnmt3a-Dnmt3L) fused to dCas9 enhanced methylation ~4–5 times in human cells compared to Dnmt3a alone (Stepper *et al.*, 2017). Other hybrid approaches for enhancing demethylation of target regions are discussed in Section 7.2 on site-specific activators and repressors.

[13] The SunTag toolbox consist of a repeating polypeptide array that can recruit multiple copies of antibodies, which can bind multiple protein catalytic domains. Coupled with dCas9, the SunTag-dCas9 complex shows enhanced epigenetic effects (Tanenbaum *et al.*, 2014).

Histone modifiers

Histones H3 and H4 are the evolutionary most conserved histones, suggesting that they play a more prominent regulatory role in chromatin formation. In eukaryotic cells, the N-terminal tails of nucleosomal histones are post-translationally modified with a variety of molecules, *e.g.* methylation (addition of a methyl group), acetylation (addition of an acetyl group) and others[14]. Some of these modifications are associated with chromatin remodelling and regulation of gene activity (Berger *et al.*, 2009; Miglani *et al.*, 2020). For instance, the chromatin structure at cis-acting enhancer[15] regions in the genome is altered by characteristic histone modifications: 'primed and poised' enhancers in the genome are marked with histone 3 methylation at lysine 4 (H3K4me1 and H3K4me2) or trimethylation at lysine 27 (H3K27me3), whereas 'active' enhancers are marked by the acetylation of the lysine 27 of histone H3 (H3K27ac) and consequently low or absent H3K27me3 (Heinz *et al.*, 2015).

There is a growing interest in the possibility to alter specific histone modifications. This can be achieved by linking a DNA binding domain to a histone-specific methylase or acetyltransferase.

Histone modifiers: detailed mode of action

Various fusion constructs have been used between a DNA binding domain and a protein acting on specific posttranslational modifications of histone proteins. TALE repeat domains have been fused to the H3 lysine-specific demethylase LSD1, which catalyses the removal of the methyl groups from H3K4 with consequent down-regulation of proximal genes (Mendenhall *et al.*, 2013). LSD1 is an anti-cancer target and its biological role was recently shown to go beyond demethylation of H3 (Gu *et al.*, 2020).

The acetyltransferase core domain of the human EP300 protein can be fused to ZFs, TALEs, and dCas9 variants to activate endogenous human genes by enrichment for H3K27ac at promoter or enhancer loci (Hilton *et al.*, 2015). The dCas9 fused with acetyl transferase (HAT) was shown to improve drought stress in *Arabidopsis in vivo* when targeted to the abscisic acid (ABA)-responsive element binding protein 1/ABRE binding factor (AREB1/ABF2) (Roca Paixão *et al.*, 2019).

Histone deacetylases (HDACs) catalyse the removal of the acetyl group from histone tails. A library of 223 chromatin regulators, including deacetylases, were identified in yeast; fusion to a zinc finger protein affected chromatin-mediated transcriptional regulation in yeast (Keung *et al.*, 2014).

Inducible modifications: The modifications induced by DNA methylation modifiers consist of the addition or removal of one methyl group to or from the cytosine at specific targets and in particular at CpG-rich sites across the promoter region of genes. Histone acetylation or methylation modifiers induce the addition or removal of acetyl (or methyl) groups from the lysine residues in the tail of H3 histones, in particular at positions 27 (H3K27) and 4 (H3K4).

Result: Techniques modifying the epigenetic state of cells have been used mainly to study epigenetic regulation and advancement in programmable epigenetic editors. Correlation of epigenome editing and changes in gene expression supports the hypothesis that chromatin modifications at regulatory sites (promoters or enhancers) are causative in the regulation of transcriptional output.

Targeted gene repression is an important future application in both applied and basic research (Siddique *et al.*, 2013). Potential therapeutic approaches have been identified in areas where activation or suppression of gene expression can lead to a health impact: down-regulation of genes for viral infection, oncogene silencing and reactivation of tumour suppressor genes, imprinting disorders, induction of cell pluripotency, genetic diseases, etc. (Kungulowski and Jeltsch, 2016).

Developing cells continuously change their epigenetic state to modulate gene expression in response to local and timely requirements. Not surprisingly, the duration of the epigenetic effect across cell generations has been shown to vary between target sites and experiments. This may relate to different experimental settings and the severity of the modulating action or to the different chromatin context targeted by the catalytic domains (Kungulowski and Jeltsch, 2016). In *Arabidopsis*, demethylation by ZFN-TET1 and SunTag-Tet1 was heritable in the absence of the transgene at one target locus, but not at another. The authors hypothesised that the heritable phenotype was due to complete demethylation at the target, while the incomplete demethylation at the other target may be restored by the endogenous chromatin modulators (Gallego-

[14] Also in prokaryotes and fungal cells histone-like proteins and methylation-like mechanisms exist (Willbanks *et al.*, 2016), but these are not described in this document.

[15] Enhancers are short region of DNA that can be bound by proteins, named transcription factors, to increase the likelihood that transcription of a particular gene will occur. Enhancers are first primed or poised before becoming active at enhancing gene transcription.

Bartolomé *et al.*, 2018). A heritable stable pattern of methylation could be induced in *Arabidopsis* by a SunTag system with the *Nicotiana tabacum* DRM methyltransferase catalytic domain (NtDRMcd) with preference for CHH (H=A,G,T) target sites. 'T4-' (without transgene) plants derived from 'T3-' plants maintained *fwa* promoter methylation and the early flowering phenotype. A positioning effect of the sgRNA-dCas9 SunTag-VP64 complex in relation to the transcription start site has been reported. When a guide 170 bp upstream from the original sgRNA was used, *fwa* was also found upregulated but to a lesser extent (Papikian *et al.*, 2019).

Stable integration in transgenic *Arabidopsis* of dCas9 fused with histone acetyl transferase ($dCas9^{HAT}$) targeted to genes regulating the transcription of other genes increased gene expression and improved drought stress tolerance (Roca Paixão *et al.*, 2019).

Target organisms: Epigenetic editors have been applied essentially to different lines of cultured human cells (Siddique *et al.*, 2013; Bernstein *et al.*, 2015; Maeder *et al.*, 2013a; Liu *et al.*, 2016; Stepper *et al.*, 2017; Hilton *et al.*, 2015; Mendenhall *et al.*, 2013), mice (Liu *et al.*, 2016), model plants like *Arabidopsis* (Gallego-Bartolomé *et al.*, 2018; Roca Paixão *et al.*, 2019) and bacterial cells (Chaikind and Ostermeier, 2014; Nomura and Barbas, 2007).

Known off-target effects: Fusion constructs with a single effector protein, such as dCas9-Dnmt3A, lack the ability to fine-tune on- versus off-target DNA methylation and exhibit extensive off-target activity. When the construct expression is increased to attain high on-target methylation, the off-target methylation rate was also found to reach high levels. Use of modular systems like the dCas9-SunTag system that can recruit multiple catalytic domains has much reduced off-target effects compared to the single fusion dCas9-Dnmt3a (Pflueger *et al.*, 2018). Typically, the effects can be widespread and have been reported up to 1,000 bp away from the target as a consequence of chromatin remodelling (Stepper *et al.*, 2017).

For reducing off-target effects, a split DNA methyltransferase approach has also been used. This approach directs methylation between the specific DNA binding sites of two zinc finger pairs, thus achieving single-site, targeted methylation (Chaikind and Ostermeier, 2014).

Limitations and gaps in knowledge: Despite much basic research, further advances in the knowledge of regulation and function of chromatin modifications and their maintenance together with an enhanced specificity of DNA recognition domains are necessary to master a fine control over epigenome editing (Kungulowski and Jeltsch, 2016). The uncertain duration of the epigenetic effect has limited the potential of this technique for durable applications.

Methylation or demethylation efficiency is sometimes inconsistent across experimental models and the use of several guide RNAs targeted at the same promoter site does not always increase the efficiency of methylation (Stepper *et al.,* 2016). However, the use of the modular SunTag system was shown to increase the efficiency of the epigenetic change. There is limited information from studies investigating the application of epigenetic changes in plants.

There are still unresolved issues around the methylation of DNA. DNA methylation occurs prominently on the fifth position of the cytosine pyrimidine ring (5-methylcytosine, 5mC) at CpG occurrences in eukaryotic genomes. Methylations on the sixth position of the adenine purine ring (N6-methyladenine, 6mA) and on the fourth position of the cytosine pyrimidine ring (N4-methylcytosine, 4mC) were believed to be minimal in eukaryotes. However, recent studies have shown the unexpected presence of 6mA in a large number of eukaryotic organisms, including fungal species, plants (*Arabidopsis*, rice, strawberry, rose, green alga), animals (fruit fly, mouse, zebrafish and pigs), and human (Liu *et al.*, 2019). Its meaning is poorly understood and requires further advances to determine relationships with gene regulation activities (Kungulowski and Jeltsch, 2016). Both types of DNA methylation could play a role in epigenetic regulation (Kumar *et al.*, 2018, Liu *et al.*, 2019). Because conflicting results have been obtained by other research groups, it remains unclear whether N^6-methyladenine is widespread across higher organisms or may be an artefact of bacterial DNA/RNA contamination (Douvlataniotis *et al.*, 2020).

Finally, out of the wide range of histone modifications, not all are considered epigenetic in nature. Many of them play a role in dynamic processes such as transcriptional induction and DNA repair (Berger *et al.*, 2009). It is, therefore, hypothesised that some epigenetic repressive histone marks (*e.g.* H3K9me3, H3K27me3) may not be causative for the repression of gene transcription (O'Geen *et al.*, 2017).

7.2 Site-specific activators and repressors (CRISPRa and CRISPRi)

Introduction: The transcription of genes in eukaryotic genomes is regulated by a large number of cis-regulatory elements such as enhancers and silencers. For instance, genome-wide profiling of DNaseI hypersensitive sites in the human genome has discovered over a million of such cell-type specific regulatory elements (Thurman *et al.*, 2012). Histone modification patterns control the binding of transcription factors to these regulatory sites, and thereby regulate gene transcription. In addition to the use of 'direct' chromatin modifiers (Section 7.1), transcription factors acting as activators or repressors may be fused to a DNA binding domain for the targeted alteration of chromatin marks at specific genomic loci (Kungulowski and Jeltsch, 2016). Many variants of such activating or repressing 'epi-effectors' have been developed in the past few years, some of them under inducible regulation factors. Such systems may have promising applications for epigenome editing, but have been used largely in research to study gene regulation processes up to now.

Mechanism: By modifying the epigenetic state of DNA and histones, transcriptional activators and repressors regulate the transcription of genes. Transcriptional activators work by promoting the recruitment of chromatin modifiers, which cause chromatin decondensation, the accumulation of histone marks such as the acetylation of histone H3 at lysine 27 (H3K27ac) and the trimethylation of histone H3 at lysine 4 (H3K4me3), and the binding of RNA polymerase II (Pol II) to initiate mRNA transcription. In contrast, transcriptional repressors work through the recruitment of inhibitory transcription factors or histone modification enzymes, such as histone methyltransferases, which increase the levels of trimethylation of histone H3 at lysine 9 (H3K9me3) and cause local compaction of chromatin (Pei *et al.*, 2020).

The fusion of a DNA binding domain, usually from CRISPR-Cas9, to an epi-effector domain creates CRISPRi and CRISPRa tools for the site-directed inhibition and activation of target genes. This can be achieved through different approaches.

Site-specific activators and repressors: detailed mode of action

A number of distinct approaches have been described to achieve gene repression. Even in the absence of a repressor domain, dCas9 alone (nuclease-deactivated or dead Cas9), co-expressed with a single-guide RNA, can sterically interfere with transcriptional elongation, RNA polymerase binding, or transcription factor binding and thereby prevent gene expression (Qi *et al.*, 2013). This approach has been called CRISPRi (CRISPR interference) and is mechanistically distinct from RNAi-based silencing, which requires the destruction of already transcribed mRNAs (Miglani *et al.*, 2020). CRISPRi can also be targeted to regulatory networks, *e.g.* by repressing transcription factor expression. This works very well in bacteria and is used in various industrial biotechnology applications (Bickard *et al.*, 2013; Cho *et al.*, 2018). Effective gene repression in mammalian cells has been demonstrated, but is not very effective (Qi *et al.*, 2013).

To increase transcriptional repression, the codon-optimised dCas9 has been fused to nuclear localisation signals (NLSs) and the (Krüppel associated box) repressor KRAB. When localised to DNA, KRAB recruits a heterochromatin-forming complex that causes histone methylation and deacetylation (Thakore *et al.*, 2015). This enhanced gene repression in human cells in a target-specific manner, showing the potency of CRISPRi (Gilbert *et al.*, 2013). Other repressors have been used successfully as well, and in a comparison among several, FOG1, which has been associated with acquisition of H3K27me3 and loss of histone acetylation, fused to both termini of dCas9, showed the strongest repression at the HER2 target locus in human cells (O'Geen *et al.*, 2017). Delivery of multiple sgRNAs with dCas9-fused transcriptional repressors (KRAB, Hairy, SID, SRDX, and ERD), however, demonstrated that the efficiency of these five repressors was not consistent with the same sgRNA (Wang *et al.*, 2019d).

A further derivative of CRISPRi is called CRISPRa, allowing the transcriptional activation of specific genes. Transcriptional activators have a DNA binding domain and a domain for activation of transcription. The activation domain can recruit general transcription factors or RNA polymerase to the gene sequence, can act to move nucleosomes on the DNA or modify histones to increase gene expression. Activators may consist of fusions between a ZF or TALE and the viral protein VP16 (Liu *et al.*, 1997) or repeats of VP16, called VP64, which generally function as more robust transcriptional activators (Beerli *et al.*, 1998; Zhang *et al.*, 2011; Hilton and Gersbach, 2015). Fusion of two VP64 domains flanking dCas9 enhanced the endogenous *Myod1* gene expression to a level sufficient for reprogramming cell phenotype of mouse embryonic fibroblasts into skeletal myocytes (Chakraborty *et al.*, 2013). An improved transcriptional regulator was obtained through the rational design of a tripartite activator, VP64-p65-Rta (dCas-VPR), fused to dCas9 (Chavez *et al.*, 2015). Also the high-fidelity version of enAs-Cas12a[16] (enAsCas12a-HF1) fused to VPR resulted consistently in a relevant

[16] enAs-Cas12a = enhanced *Acidaminococcus sp.* Cas12a

gene activation (range of 10- to 10,000-fold upregulation) using pools of crRNAs, with reduced off-target effects (Kleinstiver *et al.*, 2019).

Structure-guided design and optimisation of the transcriptional activation complex enhanced the versatility and consistency of activation for multiple sgRNAs. The attachment of an RNA aptamer[17] to the sgRNA structure allowed the dCas9 complex to bind more activator effectors, which synergistically improved transcriptional activation of various human genes (Konermann *et al.*, 2015). Cheng and collaborators (Cheng *et al.*, 2016) developed an RNA-aptamer based approach called Casilio, which is a modular and versatile platform that can append one or more PUF-binding sites to its sgRNA (sgRNA-PBS). Fusion of a transcriptional activator, *e.g.* p65HSF1, to the PUF aptamer demonstrated superior activity in respect to the dCas9-p65HSF1 fusion protein. The Casilio system was also used to increase the efficiency of demethylation from a TET1 fusion, which enhanced activation of methylation-silenced genes (Taghbalout *et al.*, 2019).

An alternative approach to recruit multiple transcriptional modulator proteins is the SunTag toolbox, a repeating polypeptide array that can recruit multiple copies of antibodies. Through attaching transcriptional factors onto the antibodies, the SunTag dCas9 activating complex amplifies its recruitment of transcriptional factors and showed increased activation of target genes (Tanenbaum *et al.*, 2014).

Further modifications towards the design of epigenetic editors include systems that respond to induced stimuli in order to offer a spatial-temporal control over the machinery for targeted control of gene activity. An on/off switch was made by fusing protein domains that dimerise in response to a chemical compound like rapamicin or doxycycline to reconstitute the fully functional complex (Holtzmann *et al.*, 2018). The use of split dCas9 activators that dimerise following chemical induction further minimises off-target effects (Pickar-Oliver *et al.*, 2019). CRISPR-GO is another chemically (abscisic acid) inducible and reversible system that can control spatial chromatin interactions by coupling the CRISPR-dCas9 system with nuclear-compartment-specific proteins via ligand-mediated dimerisation (Wang *et al.*, 2018a). In contrast to chemical induction, FACE is a far red-light activated CRISPR-dCas9 effector system based on dCas9 and the bacterial phytochrome BphS. Far-red light induces transcriptional activation of exogenous or endogenous genes with low background activity, deep tissue penetration and negligible phototoxicity (Shao *et al.*, 2018). CASANOVA combines the potent anti-crispr AcrIIA4 from *Streptococcus pyogenes* with the LOV2 photosensor from *Avena sativa*. The Acr–LOV2 hybrid showed potent Cas9 inhibition in the dark and almost full recovery of Cas9 activity after photoactivation with blue light (Bubek *et al.*, 2018).

Inducible modifications: Transcriptional activators and repressors alter the context around the target locus, thereby activating or repressing gene transcription. The site-specific binding of transcriptional modulators can provide steric hindrance for the RNA polymerase to initiate transcription, or the delivery of the epi-effectors at the target locus may change the DNA or histone modifications (*i.e.* the epigenetic state) and locally open up the chromatin. Most of the studies were designed to achieve better efficiency of transcriptional activation or repression. Addition of multiple factors promoting transcription (VPR), fused to a DNA binding domain, enhanced gene activation compared to single VP16 or multimerised VP64 (Hilton and Gersbach, 2015; Katayama *et al.*, 2019; Chavez *et al.*, 2015). Several scaffold systems have been developed for the flexible binding of multiple effectors (*e.g.* SAM, SunTag and Casilio) (Konermann *et al.*, 2015; Cheng *et al.*, 2016; Tanenbaum *et al.*, 2014).

Result: Site-directed transcriptional activation or repression results in the enhanced or reduced transcription of target genes. A dCas9 activator can create phenotypes in (transgenic) plants that are similar to those observed when a transgene is inserted for overexpression. For instance, in *Arabidopsis*, a dCas9 activator changed the number and size of leaves and made the plants better able to handle drought (Park *et al.*, 2017). Cells can be reprogrammed or differentiated from one cell type into another by increasing the expression of certain genes important for the formation or maintenance of a cell type (Chavez *et al.*, 2015).

Use of epi-effectors is considered attractive for understanding the regulation pathways of genes. It has allowed the high-resolution functional annotation of genomic regulatory elements across cell populations and developmental stages (Hilton and Gersbach, 2015). Chromatin remodellers can create loops between distant parts of chromatin in a targeted way or between genomic loci and other nuclear structures, thus allowing to scan for distal enhancers (Pickar-Oliver *et al.*, 2019; Wang *et al.*, 2018a). The availability of versatile tools for DNA recognition is exploited for high-throughput screening to annotate the non-coding genome and identify putative regulatory elements in their native chromosomal context (Klann *et al.*, 2017; Simeonov *et al.*, 2017).

[17] An RNA aptamer is an RNA oligonucleotide with specific protein-binding activity

Clinical uses of epigenetics modulators are limited. DNA methyltransferase and histone deacetylases have been used to treat certain types of tumour (Helin and Dhanak, 2013) or other inhibitory compounds in preclinical studies; however, if not coupled to a DNA binding domain, they cannot be targeted to a specific genomic region, rather they are active on a large scale (Holtzmann *et al.*, 2018). Targeted epigenomic editing holds promises to treat conditions related to aging, cell differentiation or imprinting of mutated alleles. In addition, epigenetic editing systems can be used to reprogram cell fate, *e.g.* delivery of dCas9-VP64 activated sustained expression of the myoblast determination protein Myod1, involved in regulating muscle differentiation, and diverted mouse embryonic fibroblasts into skeletal myocytes (Chakraborty *et al.*, 2013).

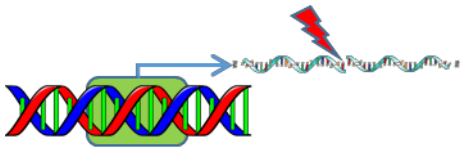
Target organisms: Epigenomic modulators have been used in human cells (Liu *et al.*, 1997; Zhang *et al.*, 2011; Gilbert *et al.*, 2013; Maeder *et al.*, 2013b; Katayama *et al.*, 2019; Chavez *et al.*, 2015; Kleinstiver *et al.*, 2019; Konermann *et al.*, 2015; Cheng *et al.*, 2016; Taghbalaout *et al.*, 2019; Tanenbaum *et al.*, 2014; Morita *et al.*, 2016; Morgan *et al.*, 2017), insects (Wang *et al.*, 2019d), mice (Katayama *et al.*, 2019; Morita *et al.*, 2016), yeast (Tanenbaum *et al.*, 2014) and plants (Gallego-Bartolomé *et al.*, 2018).

Off-target modifications: Inducing epigenetic changes or modifying the transcriptional status of targeted regions via epi-effectors fused to dCas9 may result in off-target alterations at related sites (Maeder *et al.*, 2013b). However, transcriptional activators may not necessarily affect changes in gene expression when binding to off-target sites. A recent study revealed minimal changes in gene expression induced by an RNA-guided dCas9-KRAB repression domain (Gilbert *et al.*, 2013). SunTag showed minimal effects on genome-wide methylation or gene expression (Gallego-Bartolomé *et al.*, 2018). Why certain combinations of cell type, genomic loci and epigenetic effectors are more prone to off-target effects than others remain an unsolved issue (Hilton and Gersbach, 2015).

Limitations and gaps in knowledge: Designing fusion complexes between DNA binding domains and effectors requires optimisation and fine-tuning. For instance, activator and repressor domains work when fused to the C-terminus of dCas9 (Chavez *et al.*, 2015), but may be stronger when fused to the N-terminus, whereas fusion to both termini could completely abolish any activity (O'Geen *et al.*, 2017). Similarly, it was shown that the toolbox of epigenetic editors may have profoundly different effects depending on the cell type and target locus.

A significant increase in understanding the epigenome functioning is required to bridge the gap and lead to consolidated applications of epigenome editing: *e.g.* why the same epigenetics tools induce different levels of gene expression in different cell lines and why the expression is modestly affected in many cases; what is the significance of the many epigenetic marks, their interplay and the causal relationships between change of their status and observed phenotypic variations; and, most importantly, the stability and inheritance of some epigenetic perturbations as opposite to the transient nature of others. Ongoing collaborative research efforts (ENCODE) are expected to cast light on some of these critical aspects, *e.g.* on how cell and tissues make use of the information encoded in the regulatory regions of the human and mouse genomes and how methylation changes during development in relation to changes in patterns of gene expression (Abascal *et al.*, 2020).

8 Group 4: Site-directed RNA editing



In living organisms, the information stored in the genome is translated into proteins via an intermediate compound, the messenger RNA (mRNA). The single-stranded mRNA is formed from the complimentary DNA sequence through a process called gene transcription. In recent years, it became clear that the mRNA enables an additional layer of control prior to its translation into proteins. Several tools have been developed that allow to reprogram the genetic information at the RNA level by modification of single bases (site-directed RNA base editing) or of pre-mRNA splicing reactions controlling the inclusion or exclusion of exons. In addition, techniques have been developed for the removal of specific mRNAs from a cell through RNA-specific cleavage, a process called RNA interference.

In contrast to the editing of DNA in the genome, the editing of mRNA has several advantages in certain contexts. RNA editing does not affect the DNA sequence, therefore, it can be considered a transient process that has no long-term, potentially detrimental, off-target effects. Another advantage is that it can be applied also to post-mitotic, differentiated cells.

8.1 RNA base editing

Introduction: Single RNA bases can be modified by a technique called site-directed RNA base editing. It relies on endogenous genes or engineered variants thereof, and can provide temporary solutions, *e.g.* for treating human diseases (Montiel-Gonzalez *et al.*, 2019).

'Adenosine Deaminases that Act on RNA' enzymes (ADARs) are RNA modifying enzymes with an important biological function in eukaryotes. They are structurally composed of a double stranded (ds) RNA binding domain and a deaminase domain. The catalytic domain converts adenosine (A) to inosine (I) in dsRNA by deamination. As an I is read as a G during translation, this functionally changes the mRNA codon sequence from A to G, which may be translated into amino acid changes or lead to different splicing variants (Nishikura, 2016). Recently, an engineered version of ADAR2 was created for converting cytidine (C) to uridine (U), expanding the toolbox for single nucleotide base editing of RNA (Abudayyeh *et al.*, 2019).

Mechanism: Site-directed RNA (base) editing (SDRE) uses different approaches for editing target RNA residues using ADARs (reviewed in Montiel-Gonzalez *et al.*, 2019). The challenge is to redirect ADAR activity to a target that is not already naturally deaminated and to create a structure that enables the desired editing activity by the ADAR deaminase domain. Two strategies have been employed to create editable dsRNA structures: 1) native ADAR enzymes can be recruited to a target RNA by the provision of an antisense RNA oligonucleotide (Montiel-Gonzalez *et al.*, 2019; Merkle *et al.*, 2019), and 2) the catalytic domain of an engineered ADAR can be fused to a sgRNA or to a catalytically-inactive Cas13b that is used together with a sgRNA. The sgRNA functions both to drive the engineered ADAR to the target RNA and to create an editable dsRNA structure (Cox *et al.*, 2017; Abudayyeh *et al.*, 2019).

Site-directed RNA base editing: detailed mode of action

One approach for site-directed RNA base editing is based on taking advantage of the endogenous human ADARs to achieve base conversion, obviating the need to deliver (engineered) ADAR proteins to the cell. It was called RESTORE (recruiting endogenous ADAR to specific transcripts for oligonucleotide-mediated RNA editing) and employs chemically stabilised antisense oligonucleotides (Merkle *et al.*, 2019). These engineered oligonucleotides comprise: i) a domain for the recruitment of ADARs containing several modifications and ii) a 40 nt specificity domain, that drives the ADAR to the target (Merkle *et al.*, 2019). Similar to oligonucleotide-mediated mutagenesis acting on DNA (Section 6.1), RESTORE only requires the chemically stabilised oligonucleotides to be delivered to the cells, *e.g.* by transfection. The editing yields vary from approximately 7 % to 30 % without treatment with interferon α and from 20 % to 40 % in the presence of interferon α (which increases expression of the targeted ADAR1 p150 isoform). Editing can be targeted at several transcripts simultaneously with RESTORE, by co-transfection of two antisense oligonucleotides (Merkle *et al.*, 2019).

Instead of recruiting endogenous ADARs, which are not always widely available, five different approaches have been developed for the use of engineered ADARs that are co-delivered to the treated cells. Three of these strategies differ in the method used to couple the ADAR deaminase domain to the guide RNA. Two additional approaches couple the ADAR to Cas13b instead of to the sgRNA.

The SNAP-tag couples the sgRNA with the deaminase domain of ADAR. A mutated O^6-alkylguanine-DNA alkyltransferase (hAGT) fused to the N-terminus of human ADAR1 or ADAR2 deaminase domains, catalyses the binding of a synthetic O-benzylguanine (BG)-labelled sgRNA to the SNAP-tag deaminase domain fusion protein when both are expressed in a cell (Keppler *et al.*, 2002; Stafforst *et al.*, 2012; Juillerat *et al.*, 2003). The sgRNAs are chemically stabilised 17 nt long antisense oligonucleotides, perfectly matching the target mRNA except for a single cytidine mismatch under the target adenosine, which is positioned in the middle of the sgRNA. Depending on the editor version used, editing efficiency ranged from approximately 30 % to 90 % (Vogel *et al.*, 2014 & 2018). The use of a deaminase domain with a mutation that increases the catalytic activity of ADAR, E488Q, generally increased editing efficiency. Chemical modifications added to the sgRNA to stabilise it at the same time influence editing efficiency on target and off-target adenine residues. This can be exploited to modulate the specificity and efficiency of the system (Schneider *et al.*, 2014).

In the λN-BoxB system, ADAR's deaminase domain (ADARDD) is linked to the sgRNA by using a 22 amino acids peptide from the bacteriophage Lambda N protein (λN) and its binding site is composed of 15-17 nt RNA stem-loops, called BoxB. A repeat of four λN units in tandem fused to the deaminase and of two BoxB hairpins linked to the sgRNA were found to be most effective (Montiel-Gonzalez *et al.*, 2013 & 2016; Paul *et al.*, 2002). The sgRNAs for this system are composed of: i) a central region of 29 nt complementary to the target mRNA except for a cytidine mismatch with the target adenosine, ii) two BoxB RNA hairpins and iii) two fully complementary regions of 10 nt on each terminus of the guide (Paul *et al.*, 2002). Both the sgRNA and the λN-ADARDD fusion must be transfected to the cells on an expression vector. Editing efficiency ranged from 20 % to 70 % depending on the target. The E488Q mutation increased on-target efficiencies, but the result was highly depending on the context.

Similarly to the λN-BoxB system, the bacteriophage MS2 protein also recognises RNA hairpins. Loops with AUCA sequence and an unpaired adenosine two nucleotides before the loop on the 5' arm of the stem are bound with high affinity by the protein (Keryer-Bibens *et al.*, 2008). The deaminase domain of human ADAR1 was fused downstream of MS2. The sgRNAs are composed of: i) 21 nt complementary to the target mRNA, with the target A located in the middle in the 3' region and ii) 6X tandem repeats of the stem-AUCA loop RNA sequence in the 5' region. An expression vector transfected to the cells expresses both the sgRNA and the MS2-ADARDD fusion protein. A sgRNA cytidine mismatch with the target was used to favour editing and in some cases a G mismatch was used to inhibit editing at off-target sites (Azad *et al.*, 2017).

The Cas13b RNA adenine base editor platform REPAIR was developed by fusing a catalytically-inactive Cas13b (dCas13) to an engineered human ADAR. The dCas13b domain is guided to the target RNA site by a sgRNA and both need to be delivered to the cells, *e.g.* as an expression vector (Cox *et al.*, 2017). Cas13b does not require the presence of PFS (protospacer flanking sequences) or PAM sequences and does not have a motif preference surrounding the target adenosine. Therefore, any adenosine in the transcriptome could be potentially targeted by placing a cytidine across the target adenosine in the guide RNA (Cox *et al.*, 2017). Multiplex editing should be possible with the platform, because Cas13b has pre-CRISPR-RNA (crRNA) processing activity (Smargon *et al.*, 2017). The editing efficiency of this platform over all tested targets was up to around 30 % (Cox *et al.*, 2017).

The Cas13b RNA RESCUE cytidine base editor platform uses a catalytically-inactive Cas13b from *Riemerella anatipestifer* (dCas13b) fused to a human ADAR2 that has been engineered with rational mutagenesis to catalyse C to U editing on dsRNA. The dCas13b domain is guided to the target using a 30 nt guide, and C or U base-flips across the target base are optimal for RESCUE activity. Cytidine flanked by a U or A showed to undergo robust editing activity, while sites with a 5' C or G were edited with a lower efficiency. The final RESCUE-S version of the platform provides approximately 76 % on-target C to U editing (Abudayyeh *et al.*, 2019).

Inducible modifications: Directed deamination of adenosine to inosine or of cytidine to uridine at specific RNA target sites in eukaryotic cells. Inosine is functionally equivalent to guanine in translation and in splicing (Nishikura *et al.*, 2010; Tan *et al.*, 2017), and uridine is the equivalent of thymidine, therefore, edited RNA may be translated in proteins with a changed amino acid, or in a splicing variant (see Section 8.4).

Result: RNA editing could be employed for treating diseases that originate from transient changes in gene expression, such as local inflammation. It would also be possible to correct mutations within mRNAs. In particular diseases caused by mutations that introduce termination codons such as Cystic Fibrosis (Montiel-Gonzalez *et al.*, 2013) or early onset of Parkinson Disease (Wettengel *et al.*, 2016) are possible targets. Some

strategies were also tested in the context of missense mutations, which are associated with diseases like the Factor V Leiden Thrombophilia (Vogel *et al.*, 2014) or the Rett Syndrome (Sinnamon *et al.*, 2017). The transient nature of the induced modifications may require repeated administration of the treatment. The RESTORE system was used with good to moderate yields in many primary human cells to target two mutations: the phosphotyrosine 701 in the endogenous gene STAT1 (O'Shea *et al.*, 2015) and the PiZZ mutations in SERPINA1 (serpin family A member 1), which causes severe damage to the lungs and the liver (Lomas and Mahadeva, 2002; Merkle *et al.*, 2019).

Target organisms: This technique can be applied to mammalian cells, including human cells (Merkle *et al.*, 2019; Vogel *et al.*, 2014; Vogel *et al.*, 2018; Sinnamon *et al.*, 2017; Azad *et al.*, 2017; Cox *et al.*, 2017; Abudayyeh *et al.*, 2019).

Known off-target effects: There are four types of possible off-target edits: in the target mRNA within the region complementary to the sgRNA; in the target mRNA outside the region complementary to the sgRNA; within the sgRNA; in non-target mRNAs (Montiel-Gonzalez *et al.*, 2019). Off-target editing of the sgRNA can cause further off-target editing due to the specificity of the mutated sgRNA for new mRNAs/regions. It is difficult to compare off-target editing between the systems because there has not been a systematic study comparing the strategy on similar targets and with harmonised methods. However, results for each single strategy are reported.

With the RESTORE strategy limited off-targets were detected by deep RNA sequencing. The number of off-targets slightly increased after interferon α application (3 sites out of 20,156 without, 14 with interferon). All off-target sites are engineered antisense oligonucleotide dependant, as supported by sequence alignment. Notably, 5 of the 14 off-targets with interferon showed attenuated editing, presumably due to steric blockade by the engineered antisense oligonucleotide of the natural editing site (Merkle *et al.*, 2019).

For the SNAP-tag strategy off-targets in the region complementary to the sgRNA were reduced by adding 2'methoxy or 2'fluoro modifications to the sgRNA in the Cs complementary to the off-target As. This limited off-target editing in the region to 20 %. The integration of the DNA coding for the SNAP-tag ADAR construct in the genome reduced off-target events outside the complementary region to sgRNA. The authors reported 6 and 30 off-target events with SNAP-tag DD from hADAR1 and 2, respectively, but many more with the ADAR mutants (Vogel *et al.*, 2018).

For the λN-BoxB strategy, the insertion of G mismatches under sgRNA non-targeted A's reduced off-target events in the target mRNA, within and outside the region complementary to the sgRNA. By redirecting the ADAR to the nucleus using a Nuclear Localization Signal added to the N-terminus of the enzyme, the number of off-targets events was significantly reduced. However, off-target events were still numerous, particularly in the presence of the E488Q version of the λN-ADARDD enzyme (Vallecillo-Viejo *et al.*, 2018).

For the Cas13b-RNA adenine base editor platform, transcriptome-wide sequencing identified very limited off-targets in the transcriptome. In particular the E488Q/T375G mutant had 20 off-target edit events. It seemed that also the DNA strand in RNA-DNA heteroduplexes could be targeted by ADAR (Zheng *et al.*, 2017). However, the inefficient DNA binding of Cas13b and the cytoplasmic localisation of the REPAIR complex makes targeting of RNA-DNA heteroduplexes by ADAR unlikely (Cox *et al.*, 2017). Also the RESCUE-S system produced a number of off-targets (Abudayyeh *et al.*, 2019).

Limitations and gaps in knowledge: Only A to I and C to U edits can be commonly performed so far, although other substitutions have been reported recently (Abudayyeh *et al.*, 2019). An increase in efficiency of target editing, also on difficult targets, is needed.

For the RESTORE native ADAR recruitment strategy, the codon scope is probably limited by the codon preferences of natural ADARs (Eggington *et al.*, 2011), although the authors have demonstrated the editing of three different codons. There is also space for refining the sequence and chemistry of the stabilised antisense oligonucleotides, to improve their pharmacological properties, to recruit ADARs more efficiently, and to expand the approach to other ADAR isoforms (Merkle *et al.*, 2019).

Concerning engineered ADAR strategies, the delivery of the different components to target cells (sgRNA and editors), and in particular the immunogenicity of the modified enzymes or the Cas protein may be important problems to overcome (Montiel-Gonzalez *et al.*, 2019). Optimisation of the sgRNA design and further engineering of the deaminase domain of ADARs may further reduce the number of off-target edits generated by the different RNA base editing platforms.

8.2 Oligonucleotide-mediated RNA interference

Introduction: Antisense oligonucleotides (ASOs) are small 16–20 nucleotides long, single-stranded fragments of DNA or a mixture of DNA and RNA designed to target specific pre-mRNA, mRNA or noncoding RNA sequences, to which they directly bind. ASOs can be used to cleave the target RNA, or to drive alternative splicing, depending on the chemical modification used for the ASO's backbone, and on the choice of the target sequence. Alternative splicing is described in paragraph 8.5. Cleavage of the target RNA can be used to reduce the amount of protein produced or to modulate the expression of a specific mRNA (Schoch and Miller, 2017). ASOs can also be used to hamper mRNA translation by sterically blocking the ribosomal subunits from attaching to or moving along the mRNA transcript (Dias *et al.*, 1999; Bennett *et al.*, 2010). In 1978, Zamecnik and Stephenson used a synthetic ASO to inhibit the translation of Rous sarcoma virus (RSV) RNA in a culture of chick embryo fibroblasts (Zamecnik *et al.*, 1978; Stephenson *et al.*, 1978). ASOs have since then gone through many advances in the understanding of their mechanism of action, in their synthesis and chemical modification to improve their characteristics and behaviour *in vivo*. New generations of ASOs and siRNAs, coupled with improvements in their delivery methods, have earned US FDA approvals (Olina *et al.*, 2018).

ASOs are different from small interfering RNAs (siRNAs), which are 19–21 nucleotides long, double-stranded RNA fragments that also target RNA for degradation, but only after binding to proteins of the RNA-induced silencing complex (RISC) (Olina *et al.*, 2018).

Mechanism: ASOs form ASO-RNA heteroduplexes with their target RNA. These heteroduplexes are a substrate for RNase enzymes present in the cytoplasm, which degrade the RNA in the heteroduplex. In particular, RNase H1, an enzyme that cleaves the RNA strand of RNA/DNA duplexes in a site-specific manner is recruited (Wu *et al.*, 2004; Crooke, 2017). Unmodified ASOs are easily degraded by intracellular endonucleases and exonucleases. Therefore, diverse chemical modifications have been developed to protect them against nuclease degradation, to maximize their activity and specificity and to reduce undesired effects. To improve ASOs performance, they have to be designed to target attractive sites in target RNAs and serve as optimal substrates for RNase H1 (Crooke, 2017). In addition to recruiting RNase H1 for mRNA degradation, ASOs also be used to control protein levels by inducing translation arrest due to steric hindrance (Dias *et al.*, 1999; Bennett *et al.*, 2010). Steric hindrance is directly related to the binding affinity of the ASO to its target RNA. Synthetic single-stranded oligonucleotides of 20–25 nucleotides can also be designed to bind miRNAs or pre-miRNA to prevent their interaction with mRNA, hence controlling gene expression (Catalanotto *et al.*, 2016; Crooke *et al.*, 2018).

Exogenous siRNAs associate with the Argonaute 2 (Ago 2) enzyme to form the RISC complex before interacting with the target RNA site. The RISC complex is then directed to the complementary mRNA region where the Ago 2 enzyme cleaves the target RNA (Rand *et al.*, 2005; Bumcrot *et al.*, 2006; Yu *et al.*, 2012; Lima *et al.*, 2012).

Inducible modifications: Targeted RNA cleavage of protein-coding (mRNA) and noncoding RNA molecules (*e.g.*, long non-coding RNAs); translation arrest of specific mRNAs; gene expression control through miRNA binding.

Result: This technique can be used for the regulation of gene expression at the transcript level. It has been extensively explored for clinical use, in particularly diseases that do not have alternative effective treatments. Recently one ASO for RNA interference: inotersen, and one siRNA: patisiran, have been commercialised (Rüger *et al.*, 2020). Inotersen is a chemically modified ASO that targets the 3'UTR of the mRNA of the transports thyroxine and retinol (TTR) protein, which is defective in patients with hereditary transthyretin (TTR)-mediated amyloidosis (hATTR) (Kerschen *et al.*, 2016; Keam *et al.*, 2018). Like inotersen, patisiran also leads to the degradation of TTR mRNA (Hoy *et al.*, 2018). However, while patisiran uses the RNAi pathway, inotersen induces RNA cleavage through RNase H1. There is also emerging research for the application of ASOs for prophylactic and therapeutic vaccine applications (Batista-Duharte *et al.*, 2020).

Target organisms: Animals, particularly humans.

Known off-target effects: Hybridisation-dependent toxicity can be caused by off-target effects, due to complete or partial complementary recognition of unintended transcript targets. For ASOs it has been demonstrated that off-target cleavage of sensitive sites, in particular in long pre-mRNAs can take place, and may account for cellular toxicity of ASOs (Burel *et al.*, 2016). For exogenous siRNAs competition for Ago2 binding between them and microRNAs could result in unexpected cellular effects due to changes in microRNAs activities (Liang *et al.*, 2016).

Limitations and gaps in knowledge: One of the main challenges for using ASOs and siRNA successfully as therapies for human diseases is the need to improve their tissue and cellular delivery (Kaczmarek *et al.*, 2017). Despite advances in the synthesis and understanding of the mechanism of action of the different chemical modifications that can be applied to ASOs in order to improve their stability, their delivery and

immunogenic characteristics, further studies are necessary. A limitation to most of these oligonucleotide-based therapies is their need for continuous (weekly or monthly) treatments. Their exorbitant cost is at the moment another important limiting factor to their applicability (Rüger *et al.*, 2020).

8.3 CRISPR-Cas-mediated PAM-dependent RNA interference

Introduction: CRISPR-Cas9 can be targeted to single-strand DNA (ssDNA) by providing an antisense oligonucleotide that anneals to the target ssDNA and includes a recognised PAM sequence, called a PAMmer (Sternberg *et al.*, 2014). Similarly, the use of a PAMmer can direct Cas9 to cut a single-strand RNA target (O'Connell *et al.*, 2014).

Mechanism: Cas9-mediated RNA targeting builds upon the Cas9 DNA-targeting requirements. Cas9 binding and activity on target DNA requires the recognition of a short DNA sequence on the opposite strand and next to the target sequence, known as the protospacer adjacent motif (PAM). For the targeting of RNA molecules, a short antisense oligonucleotide, called a PAMmer, can provide the PAM sequence *in trans*. The PAMmer, which needs to be delivered to the cells in addition to Cas9 and a sgRNA, binds adjacent to the sgRNA binding site on the target RNA. Interestingly, the Cas9-sgRNA complex apparently cleaves single-strand RNA substrates both when the PAM sequence is base-paired and when it is not (O'Connell *et al.*, 2014). In order to avoid contemporary DNA targeting and cleavage, the genomic DNA site corresponding to the RNA target must not have a PAM next to the target sequence. The system was termed RNA-targeting Cas9 or RCas9 (O'Connell *et al.*, 2014).

Inducible modifications: The provision of Cas9 together with a suitable sgRNA and PAMmer oligonucleotide results in the cleavage of target RNA (O'Connell *et al.*, 2014).

Result: This technique can be used for regulation of gene expression at the transcript level. In particular, this function could be important when the target RNA is located in compartments or organelles where the RNAi components are not present or active (Nelles *et al.*, 2015). The detection of endogenous RNA under physiological salt conditions with the dCas9 variant is another possibility, *e.g.* it has been used for the visualisation and tracking of specific mRNAs in stress granules in living cells (Nelles *et al.*, 2016). The diagnostic and therapeutic potential of RCas9 has been demonstrated by employing it for the visualisation and elimination of pathogenic RNA species associated with microsatellite-repeat expansion diseases in human cells (Batra *et al.*, 2017).

Target organisms: Human cells (O'Connell *et al.*, 2014; Nelles *et al.*, 2016; Batra *et al.*, 2017). *In vivo* efficacy still needs to be demonstrated. Truncated versions of RCas9 have been developed that are compatible with the limited capacity of adeno-associated viral (AAV) packaging (Batra *et al.*, 2017), which may be better for long-term expression (Naldini *et al.*, 2015).

Known off-target effects: Off-target effects of this technique have not been specifically characterised yet. Such effects may be similar as those associated with the application of CRISPR-Cas9 nucleases and of their catalytically inactive variants (see Sections 5.1.1.4 and 6.2).

Limitations and gaps in knowledge: Single-guide RNA and PAMmer sequences have to be carefully chosen in order to be able to discriminate between the target RNA and the corresponding genomic DNA. A compromise has to be sought between binding specificity and binding affinity, by varying the PAMmer's number of nucleotides upstream of the PAM sequence: 2–8 additional nucleotides seem to be optimal (O'Connell *et al.*, 2014). PAMmers also need to be modified for effective *in vivo* delivery, in order to stabilise them and prevent destruction of the target RNA by RNAse H1, which degrades RNA in RNA-DNA hybrids (Nelles *et al.*, 2015). The nature and extent of off-target effects on genomic DNA and other RNA targets should be investigated before the technique can be applied in the clinical setting for therapeutic purposes.

8.4 CRISPR-Cas-mediated PAM-independent RNA interference

Introduction: In recent years, various CRISPR-based tools that can target and cleave RNA have been identified. Type II (RCas9 and Cas9 orthologues SauCas9, CjCas9 and FnCas9), Type III (Csm and Cmr), and Type VI (Cas13) CRISPR-Cas systems have been studied and repurposed for RNA-based applications. RCas9 technology, using PAMmers, has been described in Section 8.2. This Section describes the application of type II (other than RCas9), III and VI CRISPR-Cas systems for RNA interference or gene knockdown. These RNA targeting systems work in a PAM-independent way.

Mechanism: CRISPR-Cas based effectors able to cleave RNA in a crRNA dependent manner were identified in type II, III and VI groups of the CRISPR-Cas classification system. Type II Cas9 orthologues SauCas9 and CjCas9 have been shown to directly cleave single-strand RNA without depending on the presence of a PAM sequence (Strutt *et al.*, 2018; Dugar *et al.*, 2018; Price *et al.*, 2015; Sampson *et al.*, 2013).

Type III CRISPR-Cas systems also cleave target RNA in a PAM-independent manner, but display additional unspecific cleavage of non-target DNA and RNA. DNA degradation seems to be additionally regulated by short protospacer flanking sequences (also called RNA PAM) 3' of the crRNA complementary region (Terns, 2018).

Type VI CRISPR-Cas effectors from the Cas13a and Cas13d families may or may not require a RNA PAM, and once activated by sgRNA, can cleave bystander, non-target RNAs *in vitro* and in bacteria. However, the collateral degradation of non-target RNAs has not been detected in eukaryotic cells. Cas13a and Cas13d nucleases can also process pre-crRNAs, and can therefore be used for multiplex targeting of RNA sequences (Aman *et al.*, 2018; Abudayyeh *et al.*, 2016 & 2017; Gao *et al.*, 2016; East-Seletsky *et al.*, 2016; Gootenberg *et al.*, 2017; Smargon *et al.*, 2017; Yan *et al.*, 2018; Konermann *et al.*, 2018; Aman *et al.*, 2018).

PAM-independent RNA interference: detailed mode of action

When targeted to RNA with the use of a sgRNA, type II CRISPR-Cas effectors SauCas9, CjCas9 and FnCas9 cleave RNA. This is different from the RCas9 technology, which requires the use of PAMmers (O'Connell *et al.*, 2014). Regions that form strong secondary structures were shown to be inaccessible to SauCas9 binding, preventing cleavage or repressing activity (Strutt *et al.*, 2018).

Type III (Class 1) CRISPR-Cas systems cleave both RNA and DNA. DNA cleavage occurs in the vicinity of the target sequence only if the DNA is actively transcribed (Ichikawa *et al.*, 2017; Kazlauskiene *et al.*, 2016; Elmore *et al.*, 2016; Estrella *et al.*, 2016; Han *et al.*, 2017). Upon activation by sgRNA binding, type II systems also have non-specific RNase activity. However, both non-specific degradation of single-strand DNA and of RNA are limited in time (Kazlauskiene *et al.*, 2016). More information on specific type III systems can be found in Terns (2018) and Varble and Marraffini (2019).

Type VI (Cas13) Cas13a and Cas13d families have been shown to target RNA. Cas13a effectors from different organisms use crRNA of 22 to 28 nucleotides in length to target RNA, with an optimal stem loop structure length of at least 24 nt. Single mismatches across the crRNA are tolerated, while there is a central 'seed region' that is sensitive to double mismatches (Abudayyeh *et al.*, 2016 & 2017; Gao *et al.*, 2016; East-Seletsky *et al.*, 2016; Gootenberg *et al.*, 2017; Smargon *et al.*, 2017; Yan *et al.*, 2018). The target RNA is cleaved at sites outside the crRNA binding region, and different Cas13a variants may have preferences for cutting after a certain base (Abudayyeh *et al.*, 2016; East-Seletsky *et al.*, 2016). Secondary structures in the target RNA and the presence of RNA binding proteins might influence the accessibility and thus the activity of Cas13a towards the target RNA (Aman *et al.*, 2018; Abudayyeh *et al.*, 2016 & 2017; Konermann *et al.*, 2018). Cas13d from *Ruminococcus flavefaciens* was engineered to develop the CasRx platform, targeting RNA in human cells for transcript knockdown or repression and for splice isoform manipulation (Konermann *et al.*, 2018), which is described in a dedicated Section (8.4). The CasRx platform consists of an engineered CRISPR-Cas13d effector, fused to N- and C-terminal nuclear localisation signals and to the human influenza hemagglutinin (HA) tag, and an associated guide RNA. This platform had a median knockdown efficiency of more than 96 %, compared to an RNA silencing efficiency on the same targets of 65 % and 53 % for small hairpin RNAs (shRNAs) and for dCas9 CRISPRi (targeting of transcriptional repressors to endogenous genes via dCas9), respectively (Konermann *et al.*, 2018; Gilbert *et al.*, 2013; Gilbert *et al.*, 2014; Zalatan *et al.*, 2015).

Inducible modifications: Targeted RNA cleavage of protein-coding (mRNA) and noncoding RNA molecules (*e.g.* long non-coding RNAs).

Result: Type II CRISPR-Cas systems: FnCas9 was used to target the 5' or 3' untranslated regions of the hepatitis C virus RNA genome in eukaryotic cells, inhibiting viral protein production and replication (Price *et al.*, 2015). FnCas9 could potentially also be used to target negative-sense ssRNA viruses.

Due to their complexity, type III systems have so far been used only in prokaryotes to investigate the function of genes and pathways (Liu *et al.*, 2018c; Zebec *et al.*, 2014, Ichikawa *et al.*, 2017). In addition to this, protocols have been refined to reconstitute active Type III-A and Type III-B complexes *in vitro* from purified components, which could be delivered as ribonucleoprotein to different cell types (Kazlauskiene *et al.*, 2016; Liu *et al.*, 2017b; Rouillon *et al.*, 2013; Samai *et al.*, 2015; Staals *et al.*, 2014; Tamulaitis *et al.*, 2014; Elmore *et al.*, 2016; Estrella *et al.*, 2016; Hale *et al.*, 2009 & 2012; Osawa *et al.*, 2015; Spilman *et al.*, 2013; Staals *et al.*, 2013; Taylor *et al.*, 2015).

Type VI systems: the relative simplicity of the Cas13 system has greatly facilitated their rapid development. Their small size renders them well suited for AAV delivery together with a CRISPR array, an optional effector domain, and regulatory elements (Konermann *et al.*, 2018). They have been harnessed as research tools for targeted RNA degradation and gene knockdown in human cells (Abudayyeh *et al.*, 2017; Cox *et al.*, 2017; Konermann *et al.*, 2018). In plants, Cas13a was used for RNA interference against the TuMV virus in tobacco and was shown to inhibit bacterial cells' growth (Abudayyeh *et al.*, 2016; Aman *et al.*, 2018). Cas13d was recently used to target SARS-CoV-2 and influenza viruses. This approach could degrade RNA from SARS-CoV-2 sequences and from live influenza A virus (IAV) in human lung epithelial cells and reduced H1N1 IAV load in respiratory epithelial cells (Abbot *et al.*, 2020). However, many further steps are needed before it can be used therapeutically. Future application may include the study of gene function and cellular pathways and the development of therapeutic agents for genetic diseases. Moreover, dCas13 systems have been applied for RNA base editing, described in Section 3.4.1 (Cox *et al.*, 2017), sensitive nucleic acid detection and patient diagnosis (East-Seletsky *et al.*, 2016; Gootenberg *et al.*, 2018; Gootenberg *et al.*, 2017; Myhrvold *et al.*, 2018), live cell transcript tracking and imaging (Abudayyeh *et al.*, 2017), and pre-mRNA splicing regulation, described in Section 8.4 (Konermann *et al.*, 2018).

Target organisms: Type II systems have been used for RNA targeting in bacteria and eukaryotic cells (Strutt *et al.*, 2018; Dugar *et al.*, 2018; Price *et al.*, 2015). Target organisms for type III systems have so far been archea and bacteria (Ichikawa *et al.*, 2017; Zebec *et al.*, 2014; Liu *et al.*, 2018c). Type VI systems were used to target RNA in human cells, mammalian cells, bacteria, rice protoplasts, tobacco leaves and plants (Konermann *et al.*, 2018; Abudayyeh *et al.*, 2016 & 2017; Aman *et al.*, 2018).

Known off-target effects: Off-target effects have not been investigated for type II CRISPR-Cas systems applied to RNA. Particular attention will have to be dedicated to DNA off-target editing for this technique, being that DNA is the primary target of Cas9 nucleases.

Off-target effects have to be further investigated for Type III systems. The presence of DNAse activity has to be either inactivated prior to employing the technique, or closely monitored to detect potential off-target effects on genomic DNA.

For type VI systems CasRx has no detectable off-target effects in human cells (Konermann *et al.*, 2018). The same observation was made for LwaCas13a (Abudayyeh *et al.*, 2017). However, the collateral cleavage activity of Cas13a and Cas13d nucleases should be closely monitored.

Limitations and gaps in knowledge: Type II systems: additional studies are needed to investigate the potential consequences of Cas9 mediated RNA targeting in eukaryotic cells.

Type III systems are composed by 6-7 Cas proteins, and therefore the expression of a functional complex in heterologous organisms and cell types is a challenge. This challenge might be overcome by the delivery of ribonucleoprotein complexes via electroporation or microinjection (Terns, 2018).

For type VI systems the capacity of activated Cas13-crRNA complexes to cleave non-complementary adjacent RNAs should be further investigated. The mechanisms that drive the target accessibility and efficiency of Cas13a and Cas13d systems also require further studies.

8.5 RNA splice isoform manipulation

Introduction: Splicing is a process in which segments of a pre-mRNA called introns are removed while segments called exons are joined together to form mature mRNA. Alternative splicing of mRNA, in which different exons are spliced together to form mRNA with varying sequences, is a ubiquitous phenomenon in eukaryotes, known *e.g.* from IgM processing in eukaryotes. Deregulation of splicing is involved in cancer and several neurodegenerative diseases in humans. It is generally regulated by the interaction of *cis*-acting elements in the pre-mRNA with positive or negative *trans*-acting splicing factors, which can mediate exon inclusion or exclusion.

A catalytically inactive version of CRISPR-Cas13d, engineered for expression in human cells, was repurposed to inhibit exon inclusion and exclusion, and therefore influence the proportion of different splice isoforms of a messenger RNA present in a cell (Konermann *et al.*, 2018; Du *et al.*, 2020).

Antisense oligonucleotides (ASOs) with various backbone chemical modifications have also been used to modulate exon inclusion and exclusion (Dominski & Kole, 1993; Qi *et al.*, 2005; Siva *et al.*, 2014; van der Wal *et al.*, 2017). These techniques can be applied to restore the correct splicing of an incorrectly spliced mRNA, to modulate the production of two different splice variants, or to create a novel splice variant.

Mechanism: Alternative splicing factors play important roles in tissue-specific alternative splicing regulation. They regulate alternative splicing positively or negatively in a position-dependent manner (Sun *et al.*, 2012). The RNA recognition motif of these splicing factors can be replaced by deactivated Cas proteins such as Cas13d from *Ruminococcus flavefaciens* engineered for the CasRx platform or Cas13b from *Prevotella* sp. P5-125 (Cox *et al.*, 2017). Such fusions can induce exon inclusion when targeted by sgRNAs to bind at a downstream intron, and induce exon exclusion when guided to bind within a target exon (Du *et al.*, 2020). Exon inclusion depends on the presence of both the Cas domain and the splicing effector. Exon exclusion can also be induced by targeting the splice acceptor with dCasRx alone, blocking access to the splicing machinery, although with lower efficiency compared to dCasRx fused to an effector (Konermann *et al.*, 2018; Du *et al.*, 2020). Separation of the RNA binding module from the exon splicing effector module and fusion of each of them to rapamycin-inducible domains created an inducible system (Du *et al.*, 2020; Luker *et al.*, 2004). Simultaneous targeting of multiple target positions with a CRISPR array achieved higher levels of exon skipping than individual guides alone (Konermann *et al.*, 2018). Another CRISPR array was able to simultaneously induce splicing modulation of two different targets (Du *et al.*, 2020).

ASOs can also be used to control exon inclusion or exclusion (Dominski & Kole, 1993; Qi *et al.*, 2005; Siva *et al.*, 2014; van der Wal *et al.*, 2017). The chemical modifications used for these ASOs have the function of avoiding RNase H recognition and subsequent mRNA degradation. They also increase target affinity and modify the uptake and distribution profile of the ASOs *in vivo* (Sazani *et al.*, 2002). Examples are phosphorodiamidate morpholinos (PMOs), peptide nucleic acids (PNA), locked nucleic acids (LNA), and 2′O-methyl (2′OMe) and 2′-O-methoxyethyl (MOE) ribose modifications (Kurreck, 2003).

Inducible modifications: Modulation of exon inclusion and exclusion. In contrast to CRISPR-Cas9-mediated genome editing strategies that act on splicing cis-regulatory elements (Gapinske *et al.*, 2018; Mou *et al.*, 2017; Long *et al.*, 2018), this strategy does not induce permanent changes in the genome and can be applied transiently and presumably reversibly. In addition to this, it is difficult to promote exon inclusion with other CRISPR-mediated DNA deletion or editing systems.

Result: Alteration of alternative splicing patterns and expression of the resulting protein isoforms. The technique was applied to a neuronal model of frontotemporal dementia and to a gene involved in spinal muscular atrophy in patient fibroblasts. In the frontotemporal dementia model dCasRx-mediated exon exclusion was able to lower the inclusion rate of the target exon to a level similar to the one present in unaffected control neurons (Konermann *et al.*, 2018). A CasRx fusion with RBFOX1 or RBM38 splicing factors (named CASFx by the authors) was able to modulate splicing of the target exon in spinal muscular atrophy patient fibroblasts (Du *et al.*, 2020). This approach offers a promising therapeutic solution for various RNA mis-splicing diseases. Exon skipping by ASOs has recently gained recognition as therapeutic approach in Duchenne muscular dystrophy (Tsoumpra *et al*, 2019). ASOs mediated exon-inclusion has proven effective *in vitro* and *in vivo* as a possible therapy for spinal muscular atrophy (SMA) (Hua *et al.*, 2012; Douglas *et al.*, 2013).

Target organisms: Human cells (Konermann *et al.*, 2018; Du *et al.*, 2020) or whole organisms (Finkel *et al.*, 2019; Mercuri *et al.*, 2018) and mice (Svasti *et al.*, 2009; Passini *et al.*, 2011).

Known off-target effects: A limited number of off-target splicing changes were identified for the CASFx platform (Du *et al.*, 2020), which may have resulted from mismatch tolerance between the sgRNA and the target RNA.

Limitations and gaps in knowledge: For Cas13-based applications, additional studies should clarify the relation between guide RNA positioning and exon skipping potential. Therapeutic uses of this technique depend on the availability of suitable delivery systems for either transient or stable expression of the splicing factors. Additional control through small molecule inducible systems may be advantageous. For ASOs, backbone chemistries are continuously developed and assessed with the purpose of improving affinity, and stability in circulation and in target cells, of enhancing cell penetration and nuclear accumulation, and contemporaneously improving resistance to nucleases (Saleh *et al.*, 2012).

9 Conclusions

This study provides an overview of the major NGTs that have been developed in plant, animal (including human) and microbial fields. Based on a systematic literature review a wide variety of NGTs with diverse mechanisms of action and applicability have been identified. A number of conclusions arise from this review.

- Particularly after the discovery of the CRISPR-Cas technology in 2012, the number of distinct NGTs developed on the basis of this technology has strongly increased. The fast pace of developments in this field also means that the full spectrum of NGTs is a moving target for reviewing as they will further evolve in the coming years. In line with the dynamic nature of the field, some of the NGTs described in this study are still in an exploratory phase and insufficiently mature to be considered sufficiently reproducible and applicable to diverse targets and different organisms.

- Despite the diversity of NGTs identified in this study we have been able to classify them in 4 groups and we believe this will be a suitable grouping for the further addition of NGTs. Whereas the first two NGT groups target the DNA sequence, as progress evolved, the focus of the techniques has been expanded to the epigenome (Group 3 NGTs) and to RNA transcribed from the DNA (Group 4). The latter two groups of NGTs provide alternative approaches for achieving certain effects without changing the heritable DNA sequence itself, and are, therefore, also included in this review.

- The NGTs developed during the past 20 years rely on molecular techniques altering the genome of individual cells, rather than being based on sexual hybridisation between plants to generate genomic modifications in progeny. The tools for generating a genome alteration, therefore, consist of DNA, RNA, protein or combinations of them. The delivery systems for the active components required by the NGT are often similar to those used for transgene delivery in genetic transformation experiments. Stable integration of the transgenes into the host genome is, however, in most cases not a prerequisite, and alternative approaches to DNA delivery (*i.e.* RNA and/or protein) may be equally effective for inducing genome alterations. Some other NGTs require the administration of only a short (DNA or RNA) oligonucleotide to the targeted cells to obtain a short genome edit. The need for different types of active components and delivery approaches is in line with the diversity of NGTs identified.

- NGT-targeted alterations are more and more precise, both in terms of being localised to a specific target site and in terms of the specific, intended DNA alteration. Compared to EGTs the alterations are generally more subtle, although insertions of long sequences may be achieved by some of the NGTs when used in combination with a suitable donor template. Consequently, the products obtained by NGTs, by hybridisation techniques or occurring naturally are becoming more and more indistinguishable from each other.

- Moreover, similar genome alterations, *e.g.* a single nucleotide substitution, can often be generated by different NGTs and a specific NGT can induce different types of genome alterations. Therefore, the technique itself cannot always be directly linked with the type of alteration that could be obtained.

- The efficiency of creating a desired genomic alteration has to be balanced against the probability of generating in parallel also unintended effects at off-target sites. Diverse optimisation strategies are employed for enhancing the specificity of the technique and for minimising off-target effects. Because the targeted sequence is known, the probability of such off-target effects could be predicted in some cases via bioinformatics analyses and then experimentally assessed. For some species, individual organisms can be selected that do not contain off-target changes, or the unintended modification may be removed in a subsequent generation by sexual crossing.

- NGTs allow to create genome alterations directly in elite germplasm or differentiated cells and thus shorten the development time for organisms with desired phenotypes and for cells to be used for gene therapy. As the changes are often small and often instructed by similar changes identified in other organisms, the resulting products containing the genome alterations display more predictable phenotypes and need less time for further testing. It may thus be expected that the technology will be increasingly deployed across the various biological kingdoms.

10 References

3,000 Rice genomes project (2014) GigaScience 3, 7 (doi.org/10.1186/2047-217X-3-7).

Abascal, F., Acosta, R., Addleman, N.J., Weng, Z. (2020) Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature 583, 699–710.

Abbott, T.R., Dhamdhere, G., Liu, Y., Lin, X., Goudy, L., Zeng, L., Chemparathy, A., Chmura, S., Heaton, N.S., Debs, R., Pande, T., Endy, D., La Russa, M.F., Lewis, D.B., Qi, L.S. (2020) Development of CRISPR as an antiviral strategy to combat SARS-CoV-2 and influenza. Cell 181, 865–876.

Abudayyeh, O.O., Gootenberg, J.S., Konermann, S., Joung, J., Slaymaker, I.M., Cox, D.B.T., Shmakov, S., Makarova, K.S., Semenova, E., Minakhin, L., Severinov, K., Regev, A., Lander, E.S., Koonin, E.V., Zhang, F. (2016) C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. Science 353, aaf5573 (doi:10.1126/science.aaf5573).

Abudayyeh, O.O., Gootenberg, J.S., Essletzbichler, P., Han, S., Joung, J., Belanto, J.J., Verdine, V., Cox, D.B.T., Kellner, M.J., Regev, A., Lander, E.S., Voytas, D.F., Ting, A.Y., Zhang, F. (2017) RNA targeting with CRISPR-Cas13. Nature 550, 280-284.

Abudayyeh, O.O., Gootenberg, J.S., Franklin, B., Koob, J., Kellner, M.J., Ladha, A., Joung, J., Kirchgatterer, P., Cox, D.B.T., Zhang, F. (2019) A cytosine deaminase for programmable single-base RNA editing. Science 365, 382-386.

Ackerman, C.M., Myhrvold, C., Thakku, S.G., Freije, C.A., Metsky, H.C., Yang, D.K., Ye, S.H., Boehm, C.K., Kosoko-Thoroddsen, T.-S. F., Kehe, J., Nguyen, T.G., Carter, A., Kulesa, A., Barnes, J.R., Dugan, V.G., Hung, D.T., Blainey, P.C., Sabeti, P.C. (2020) Massively multiplexed nucleic acid detection using Cas13. Nature 582, 277-282.

Adli, M. (2018) The CRISPR tool kit for genome editing and beyond. Nat. Commun. 9, 1911.

Ahloowalia, B.S., Maluszynski, M., Nichterlein, K. (2004) Global impact of mutation-derived varieties. Euphytica 135, 187–204.

Ain, Q.U.I., Chung, J.G., Kim, Y.H. (2015) Current and future delivery systems for engineered nucleases: ZFN, TALEN and RGEN. J. Control. Rel. 10, 120-127.

Ali, Z., Shami, A., Sedeek, K., Kamel, R., Alhasi, A., Tehseen, M., Hassan, N., Butt, H., Kababji, A., Hamdan, S.M., Mahfouz, M.M. (2020) Fusion of the Cas9 endonuclease and the VirD2 relaxase facilitates homology-directed repair for precise genome engineering in rice. Commun. Biol. 3, 44 (doi: 10.1038/s42003-020-0768-9).

Alok, A., Sandhya, D., Jogam, P., Rodrigues, V., Bhati, K.K., Sharma, H., Kumar, J. (2020) The rise of the CRISPR/Cpf1 system for efficient genome editing in plants. Front. Plant Sci. 11, 264 (doi: 10.3389/fpls.2020.00264).

Alonge, M., Wang, X., Benoit, M., ..., Lipmann, Z.B. (2020) Major impacts of widespread structural variation on gene expression and crop improvement in tomato. Cell 182, 162-176.

Aman, R., Ali, Z., Butt, H., Mahas, A., Aljedaani, F., Khan, M.Z., Ding, S., Mahfouz, M. (2018) RNA virus interference via CRISPR/Cas13a system in plants. Genome Biol. 19, 1 (doi:10.1186/s13059-017-1381-1).

Anderson, J.E., Michno, J.-M., Kono, T.J.Y., Stec, A.O., Campbell, B.W., Curtin, S.J., Stupar, R.M. (2016) Genomic variation and DNA repair associated with soybean transgenesis: a comparison to cultivars and mutagenized plants. BMC Biotechnol. 16, 41 (doi: 10.1186/s12896-016-0271-z).

Anzalone, A.V., Randolph, P.B., Davis, J.R., Sousa, A.A., Koblan, L.W., Levy, J.M., Chen, P.J., Wilson, C., Newby, G.A., Raguram, A., Liu, D.R. (2019) Search-and-replace genome editing without double-strand breaks or donor DNA. Nature 576, 149-156.

Aparicio, T., Nyerges, A., Martínez-García, E., Lorenzo, V. (2020) High-efficiency multi-site genomic editing of *Pseudomonas putida* through thermoinducible ssDNA recombineering. iScience 23, 100946 (doi: 10.1016/j.isci.2020.100946).

Arnould, C., Legube, G. (2020) The secret life of chromosome loops upon DNA double-strand break. J. Mol. Biol. 432, 724-736.

Ashikawa, I. (2001) Gene-associated CpG islands in plants as revealed by analyses of genomic sequences. Plant J. 26, 617-625.

Auton, A., Abecasis, G., Altshuler, D., *et al.* (2015) A global reference for human genetic variation. Nature 526, 68–74.

Azad, M.T.A., Bhakta, S., Tsukahara, T. (2017) Site-directed RNA editing by adenosine deaminase acting on RNA for correction of the genetic code in gene therapy. Gene Ther. 24, 779.

Aylon, Y., Kupiec, M. (2004) DSB repair: The yeast paradigm. DNA Repair 3, 797-815.

Bae, S.-J., Park, B.G., Kim, B.-G., Hahn, J.-S. (2020) Multiplex gene disruption by targeted base editing of *Yarrowia lipolytica* genome using cytidine deaminase combined with the CRISPR/Cas9 system. Biotechnol. J. 15, e1900238 (doi: 10.1002/biot.201900238).

Barrangou, R. (2015) Diversity of CRISPR-Cas immune systems and molecular machines. Genome Biol. 16, 247.

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

Batista-Duharte, A., Sendra, L., Herrero, M.J., Téllez-Martínez, D., Carlos, I.Z., Aliño, S.F. (2020) Progress in the use of antisense oligonucleotides for vaccine improvement. Biomolecules 10, 316 (doi:10.3390/biom10020316).

Batra, R., Nelles, D.A., Pirie, E., ..., Yeo, G.W. (2017) Elimination of toxic microsatellite repeat expansion RNA by RNA-targeting Cas9. Cell 170, 899-912.e10. (doi: 10.1016/j.cell.2017.07.010).

Beerli, R.R., Segal, D.J., Dreier, B., Barbas, C.F. III. (1998) Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. Proc. Natl. Acad. Sci. 95, 14628–14633.

Behler, J., Vijay, D., Hess, W.R., Akhtar, M.K. (2018) CRISPR-based technologies for metabolic engineering in cyanobacteria. Trends Biotechnol. 36, 969-1010.

Bennett, C.F., Swayze, E.E. (2010) RNA targeting therapeutics: Molecular mechanisms of antisense oligonucleotides as a therapeutic platform. Annu. Rev. Pharmacol. Toxicol. 50, 259–293.

Berg, J. (1988) Proposed structure for the zinc-binding domains from transcription factor IIIA and related proteins. Proc. Natl. Acad. Sci. U.S.A. 85, 99-102.

Berger, S.L., Kouzarides, T., Shiekhattar, R., Shilatifard, A. (2009) An operational definition of epigenetics. Genes Dev. 23, 781–783.

Bergman, Y., Cedar, H. (2013) DNA methylation dynamics in health and disease. Nat. Struct. Mol. Biol. 20, 274–281.

Bernstein, D.L., Ruano, *E.G.*, Kastner, K.H. (2015) TALE-mediated epigenetic suppression of CDKN2A increases replication in human fibroblasts. J. Clin. Invest. 125, 1998-2006.

Beurdeley, M., Bietz, F., Li, J., Thomas, S., Stoddard, T., Juillerat, A., Zhang, F., Voytas, D.F., Duchateau, P., Silva, G.H. (2013) Compact designer TALENs for efficient genome engineering. Nat. Commun. 4, 1762.

Bhargava, R., Onyango, D.O., Stark, J.M. (2016) Regulation of single strand annealing and its role in genome maintenance. Trends Genet. 32, 566-575.

Bi, Y., Hua, Z., Ren, H., Zhang, L., Xiao, H., Liu, X., Hua, W., Mei, S., Molenaar, A., Laible, G., Zheng, X. (2018) ØC31 integrase-mediated isolation and characterization of novel safe harbors for transgene expression in the pig genome. Int. J. Mol. Sci. 19, 149 (doi: 10.3390/ijms19010149).

Bickard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F., Marraffini, L.A. (2013) Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. Nucl. Ac. Res. 41, 7429-7437.

Bilichak, A., Sastry-Dent, L., Sriram, S., Simpson, M., Samuel, P., Webb, S., Jiang, F., Eudes, F. (2020) Genome editing in wheat microspores and haploid embryos mediated by delivery of ZFN proteins and cell-penetrating peptide complexes. Plant Biotechnol. J. 18, 1307-1316.

Billon, P., Bryant, E.E., Joseph, S.A., Nambiar, T.S., Hayward, S.B., Rothstein, R. Ciccia, A. (2017) CRISPR-mediated base editing enables efficient disruption of eukaryotic genes through induction of STOP codons. Mol. Cell 67, 1068-1079. e4.

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

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

Bonawitz, N.D., Ainley, W.M., Itaya, A., Chennareddy, S.R., Cicak, T., Effinger, K., Jiang, K., Mall, T.K., Marri, P.R., Samuel, J.P., Sardesai, N., Simpson, M., Folkerts, O., Sarria, R., Webb, S.R., Gonzalez, D.O., Simmonds, D.H., Pareddy, D.R. (2019) Zinc finger nuclease-mediated targeting of multiple transgenes to an endogenous soybean genomic locus via non-homologous end joining. Plant Biotechnol. J. 17, 750-761.

Breyer, D., Herman, P., Brandenburger, A., Gheysen, G., Remaut, E., Soumillion, P., Van Doorsslaere, J., Custers, R., Pauwels, K., Sneyers, M., Reheul, D. (2009) Genetic modification through oligonucleotide-mediated mutagenesis. A GMO regulatory challenge? Environ. Biosafety Res. 8, 57-64.

Brinkman, E.K., Chen, T., de Haas, M., Holland, H.A., Akhtar, W., van Steensel, B. (2018) Kinetics and fidelity of the repair of Cas9-induced double-strand DNA breaks. Mol. Cell 70, 801-813.

Broeders, M., Herrero-Hernandez, P., Ernst, M.P.T., van der Ploeg, A.T., Pijnappel, W.W.M.P. (2020) Sharpening the molecular scissors: Advances in gene-editing technology. iScience 23, 100789.

Broothaerts, W., Mitchell, H.J, Weir, B., Kaines, S., Smith, L.M.A., Yang, W., Mayer, J.E., Roa-Rodríguez, C., Jefferson, R.A. (2005) Gene transfer to plants by diverse species of bacteria. Nature 433, 629-633.

Bubek, F., Hoffmann, M., Harteveld, Z., Aschenbrenner, S., Bietz, A., Waldhauer, M.C., Börner, K., Fakhiri, J., Schmelas, C., Dietz, L., Grimm, D., Correia, B.E., Eils, R., Niopek, D. (2018) Engineered anti-CRISPR proteins for optogenetic control of CRISPR–Cas9. Nat. Meth. 15, 924-927.

Bumcrot, D., Manoharan, M., Koteliansky, V., Sah, D.W.Y. (2006) RNAi therapeutics: A potential new class of pharmaceutical drugs. Nat. Chem. Biol. 2, 711–719.

Burel, S.A., Hart, C.E., Cauntay, P., ..., Henry, S.P. (2016) Hepatotoxicity of high affinity gapmer antisense oligonucleotides is mediated by RNase H1 dependent promiscuous reduction of very long pre-mRNA transcripts. Nucl. Ac. Res. 44, 2093–2109.

Cabaniols, J.P., Pâques, F. (2008) Robust cell line development using meganucleases. In: Davis G.D., Kayser K.J. (eds) Chromosomal mutagenesis. Meth. Mol. Biol. 435. Humana Press.

Canny, M.D., Moatti, N., Wan, L.C.K., ..., Durocher, D. (2018) Inhibition of 53BP1 favours homology-dependent DNA repair and increases CRISPR–Cas9 genome-editing efficiency. Nat. Biotechnol. 36, 95-102.

Cao, L., Gao, C.H., Zhu, J., Zhao, L., Wu, Q., Li, M., Sun, B. (2016) Identification and functional study of type III-A CRISPR-Cas systems in clinical isolates of *Staphylococcus aureus*. Int. J. Med. Microbiol. 306, 686–696.

Carey, N. (2012) The epigenetic revolution: How modern biology is rewriting our understanding of genetics, disease and inheritance. New York: Columbia Univ. Press (doi:10.7312/care16116).

Carlson, D.F., Tan, W., Lillico, S.G., Stverakova, D., Proudfoot, C., Christian, M., Voytas, D.F., Long, C.R., Whitelaw, C.B., Fahrenkrug, S.C. (2012) Efficient TALEN-mediated gene knockout in livestock. Proc. Natl. Acad. Sci. U.S.A. 109, 17382-17387.

Carpenter, M.A., Law, E.K., Serebrenik, A., Brown, W.L., Harris, R.S. (2019) A lentivirus-based system for Cas9/gRNA expression and subsequent removal by Cre-mediated recombination. Methods 156, 79-84.

Carroll, D. (2014) Genome engineering with targetable nucleases. Annu. Rev. Biochem. 83, 409-439.

Casini, A., Olivieri, M., Petris, G., Montagna, C., Reginato, G., Maule, G., Lorenzin, F., Prandi, D., Romanel, A., Demichelis, F., Inga, A., Cereseto, A. (2018) A highly specific SpCas9 variant is identified by *in vivo* screening in yeast. Nat. Biotechnol. 36, 265-271.

Catalanotto, C., Cogoni, C., Zardo, G. (2016) MicroRNA in control of gene expression: An overview of nuclear functions. Int. J. Mol. Sci. 17, 1712 (doi:10.3390/ijms17101712).

Ceccaldi, R., Rondinelli, B., D'Andrea, A.D. (2016) Repair pathway choices and consequences at the double-strand break. Trends Cell Biol. 26, 52-64.

Čermák, T., Baltes, N.J., Čegan, R., Zhang, Y., Voytas, D.F. (2015) High-frequency, precise modification of the tomato genome. Genome Biol. 16, 1-14.

Čermák, T., Doyle, E.L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J.A., Somia, N.V., Bogdanove, A.J., Voytas, D.F. (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucl. Ac. Res. 39, e82.

Chaikind, B., Ostermeier, M. (2014) Directed evolution of improved zinc finger methyltransferases. PLoS One 9, 1-8.

Chakraborty, S., Ji, H.J., Kabadi, A.M., Gersbach, C.A., Christoforou, N., Leong, K.W. (2013) A CRISPR/Cas9-based system for reprogramming cell lineage specification. Stem Cell Rep. 3, 940-947.

Chalberg, T.W., Portlock, J.L., Olivares, E.C., Thyagarajan, B., Kirby, P.J., Hillman, R.T., Hoelters, J., Calos, M.P. (2006) Integration specificity of phage phiC31 integrase in the human genome. J. Mol. Biol. 357, 28-48.

Chan, M.M., Smith, Z.D., Grosswendt, S., ..., Weissman, J.S. (2019) Molecular recording of mammalian embryogenesis. Nature 570, 77-82.

Chavez, A., Scheiman, J., Vora, S., Pruitt, B.W., ..., Church, G.M. (2015) Highly efficient Cas9-mediated transcriptional programming. Nat. Meth. 12, 326-328.

Chen, F., Ding, X., Feng, Y., Seebeck, T., Jiang, Y., Davis, G.D. (2017) Targeted activation of diverse CRISPR-Cas systems for mammalian genome editing via proximal CRISPR targeting. Nat. Commun. 8, 14958.

Chen, H., Kazemier, H.G., de Groote, M.L., Ruiters, M.H.J., Xu, G.L., Rots, M.G. (2014) Induced DNA demethylation by targeting ten-eleven translocation 2 to the human ICAM-1 promoter. Nucl. Ac. Res. 42, 1563–1574.

Chen, S.P, Wang, H.H. (2019) An engineered Cas-transposon system for programmable and site-directed DNA transpositions. CRISPR J. 2, 376-394.

Cheng, A.W., Jillette, N., Lee, P., Plaskon, D., Fujiwara, Y., Wang, W., Taghbalout, A., Wang, H. (2016) Casilio: a versatile CRISPR-Cas9-Pumilio hybrid for gene regulation and genomic labelling. Cell Res. 26, 254-257.

Cheng, S., Melkonian, M., Smith, S.A., ..., Wong, G.K.-S. (2018) 10KP: A phylodiverse genome sequencing plan. GigaScience 7, 1–9.

Chevalier, B.S., Stoddard, B.L. (2001) Homing endonucleases: Structural and functional insight into the catalysts of intron/intein mobility. Nucl. Ac. Res. 29, 183757-183774.

Cho, S., Shin, J., Cho, B.-K. (2018) Applications of CRISPR/Cas system to bacterial metabolic engineering. Int. J. Mol. Sci. 19, 1089 (doi: 10.3390/ijms19041089).

Christian M., Čermák, T., Doyle, E.L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A.J., Voytas, D.F. (2010) Targeting DNA double-strand breaks with TAL effector nucleases. Genetics 186, 757-761.

Clasen, B.M., Stoddard, T.J., Luo, S., Demorest, Z.L., Li, J., Cedrone, F., Tibebu, R., Davison, S., Ray, E.E., Daulhac, A., Coffman,A., Yabandith, A., Retterath, A., Haun, W., Baltes, N.J., Mathis, L., Voytas, D.F., Zhang, F. (2016) Improving cold storage and processing traits in potato through targeted gene knockout. Plant Biotechnol. J. 14, 169-176.

Cohen, S.N., Chang, A.C., Boyer, H.W., Helling, R.B. (1973) Construction of biologically functional bacterial plasmids *in vitro*. Proc. Natl. Acad. Sci. U.S.A. 70, 3240-3244.

Collier, R., Thomson, J.G., Thilmony, R. (2018) A versatile and robust *Agrobacterium* based gene stacking system generates high quality transgenic plants. Plant J. 95, 573-583.

Corts, A.D., Thomason, L.C., Gill, R.T., Gralnick, J.A. (2018) A new recombineering system for precise genome-editing in *Shewanella oneidensis* strain MR-1 using single-stranded oligonucleotides. Sci. Rep. 9, 39, 10 pp.

Courtier-Orgogozo, V., Martin, A. (2020) The coding loci of evolution and domestication: Current knowledge and implications for bio-inspired genome editing. J. Exp. Biol. 223, jeb208934.

Cox, D.B.T., Gootenberg, J.S., Abudayyeh, O.O., Franklin, B., Kellner, M.J., Joung, J., Zhang, F. (2017) RNA editing with CRISPR-Cas13. Science 358, 1019-1027.

Crooke, S.T. (2017) Molecular mechanisms of antisense oligonucleotides. Nucl. Ac. Ther. 27, 70-77.

Crooke, S.T., Witztum, J.L., Bennett, C.F., Baker, B.F. (2018) RNA-targeted therapeutics. Cell Metab. 27, 714–739.

Cui, Y., Liao, X., Peng, S., Tang, T., Huang, C., Yang, C. (2020) OffScan: A universal and fast CRISPR off-target sites detection tool. BMC Genomics 21, 872.

Danner, E., Bashir, S., Yumlu, S., Wurst, W., Wefers, B., Kühn, R. (2017) Control of gene editing by manipulation of DNA repair mechanisms. Mamm. Genome 28, 262-274.

Davis, L., Maizels, N. (2011) DNA nicks promote efficient and safe targeted gene correction. PLoS One, 6, e23981.

Dekker, M., Brouwers, C., Aarts, M., van der Torre, J., de Vries, S., van de Vrugt, H., te Riele, H. (2006) Effective oligonucleotide-mediated gene disruption in ES cells lacking the mismatch repair protein MSH3. Gene Ther. 13, 686-694.

Delacôte, F., Perez, C., Guyot, V., Duhamel, M., Rochon, C., Ollivier, N., Macmaster, R., Silva, G.H., Pâques, F., Daboussi, F., Duchateau, P. (2013) High frequency targeted mutagenesis using engineered endonucleases and DNA-end processing enzymes. PLoS One 8, e53217.

Dermen, H. (1951) Tetraploid and diploid adventitious shoots: From a giant sport of McIntosh apple. J. Heredity 42, 145–149.

D'Halluin, K., Vanderstraeten, C., Van Hulle, J., Rosolowska, J., Van Den Brande, I., Pennewaert, A., D'Hont, K., Bossut, M., Jantz, D., Ruiter, R., Broadhvest, J. (2013) Targeted molecular trait stacking in cotton through targeted double-strand break induction. Plant Biotechnol. J. 11, 933-941.

Dias, N., Dheur, S., Nielsen, P.E., Gryaznov, S., Van Aerschot, A., Herdewijn, P., Hélène, C., Saison-Behmoaras, T.E. (1999) Antisense PNA tridecamers targeted to the coding region of Ha-ras mRNA arrest polypeptide chain elongation. J. Mol. Biol. 294, 403–416.

Ding, X., Seebeck, T., Feng, Y., Jiang, Y., Davis, G.D., Chen, F. (2019) Improving CRISPR-Cas9 genome editing efficiency by fusion with chromatin-modulating peptides. CRISPR J. 2, 51-63.

Dolgin, E. (2020) The kill-switch for CRISPR that could make gene-editing safer. Nature 577, 308-310.

Doll, N.M., Gilles, L.M., Gérentes, M.F., Richard, C., Just, J., Fierlej, Y., Borrelli, V.M.G., Gendrot, G., Ingram, G.C., Rogowsky, P.M., Widiez, T. (2019) Single and multiple gene knockouts by CRISPR-Cas9 in maize. Plant Cell Rep. 38, 487-501.

Dominski, Z., Kole, R. (1993) Restoration of correct splicing in thalassemic pre-mRNA by antisense oligonucleotides. Proc. Natl. Acad. Sci. U.S.A. 90, 8673–8677.

Douglas, A.G., Wood, M.J. (2013) Splicing therapy for neuromuscular disease. Mol. Cell Neurosci. 56, 169-185.

Douvlataniotis, K., Bensberg, M., Lentini, A., Gylemo, B., Nestor, C.E. (2020) No evidence for DNA N6-methyladenine in mammals. Science Adv. 6, 12, eaay3335 (doi: 10.1126/sciadv.aay3335).

Du, M., Jillette, N., Zhu, J.J., Li, S., Wu, A. (2020) CRISPR artificial splicing factors. Nat. Commun. 11, 2973 (doi.org/10.1038/s41467-020-16806-4).

Dugar, G., Leenay, R.T., Eisenbart, S.K., Bischler, T., Aul, B.U., Beisel, C.L., Sharma, C.M. (2018) CRISPR RNA-dependent binding and cleavage of endogenous RNAs by the *Campylobacter jejuni* Cas9. Mol. Cell 69, 893–905.

Durai, S., Mani, M., Kandavelou, K., Wu, J., Porteus, M.H., Chandrasegaran, S. (2005) Zinc finger nucleases: Custom-designed molecular scissors for genome engineering of plant and mammalian cells. Nucl. Ac. Res. 33, 5978-5990.

Dyda, F., Chandler, M., Hickman, A.B. (2012) The emerging diversity of transpososome architectures. Q. Rev. Biophys. 45, 493–521.

East-Seletsky, A., O'Connell, M.R., Knight, S.C., Burstein, D., Jamie, H.D., Cate, J.H.D., Tjian, R., Doudna, J.A. (2016) Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. Nature 538, 270–273.

Economos, N.G., Oyaghire, S., Quijano, E., Ricciardi, A.S., Saltzman, W.M., Glazer, P.M. (2020) Peptide nucleic acids and gene editing: Perspectives on structure and repair. Molecules 25, 735 (doi:10.3390/molecules25030735).

EFSA GMO Panel (2012a) Scientific opinion addressing the safety assessment of plants developed through cisgenesis and intragenesis. EFSA J. 10, 2561, 33 pp.

EFSA GMO Panel (2012b) Scientific opinion addressing the safety assessment of plants developed using Zinc Finger Nuclease 3 and other site-directed nucleases with similar function. EFSA J. 10, 2943, 31 pp.

Eggington, J.M., Greene, T., Bass, B.L. (2011) Predicting sites of ADAR editing in double-stranded RNA. Nat. Commun. 2, 319.

Elmore, J.R., Sheppard, N.F., Ramia, N., Deighan, T., Li, H., Terns, R.M., Terns, M.P. (2016) Bipartite recognition of target RNAs activates DNA cleavage by the Type III-B CRISPR-Cas system. Genes Dev. 30, 447–459.

Endo, M., Mikami, M., Endo, A., Kaya, H., Itoh, T., Nishimasu, H., Nureki, O., Toki, S. (2019) Genome editing in plants by engineered CRISPR–Cas9 recognizing NG PAM. Nat. Plants 5, 14-17.

ENGL (European Network of GMO Laboratories) (2019) Detection of food and feed plant products obtained by new mutagenesis techniques, 26 March 2019, 17 pp. (JRC116289).

Espinoza, C., Schlechter, R., Herrera, D., Torres, E., Serrano, A., Medina, C., Arce-Johnson, P. (2013) Cisgenesis and intragenesis: New tools for improving crops. Biol. Res. 46, 323-331.

Estrella, M.A., Kuo, F.T., Bailey, S. (2016) RNA-activated DNA cleavage by the Type III-B CRISPR-Cas effector complex. Genes Dev. 30, 460–470.

Fauser, F., Schiml, S., Puchta, H. (2014) Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in *Arabidopsis thaliana*. Plant J. 79, 348-359.

Feschotte, C., Pritham, E.J. (2007) DNA transposons and the evolution of eukaryotic genomes. Annu. Rev. Genet. 41, 331–368.

Finkel, R.S., Mercuri, E., Darras, B.T., …, De Vivo, D.C. (2017) Nusinersen versus sham control in infantile-onset spinal muscular atrophy. New Engl. J. Med. 377, 1723–1732.

Fitzgerald, D.M., Rosenberg, S.M. (2019) What is mutation? A chapter in the series: How microbes "jeopardize" the modern synthesis. PLoS Genet. 15, e1007995.

Fofana, B., Ghose, K., Somalraju, A., McCallum, J., Main, D., Deyholos, M.K., Rowland, G.G., Cloutier, S. (2017) Induced mutagenesis in UGT74S1 gene leads to stable new flax lines with altered secoisolariciresinol diglucoside (SDG) profiles. Front. Plant Sci. 8, 1638 (doi.org/10.3389/fpls.2017.01638).

Foster, K., Kalter, J., Woodside, W., Terns, R.M., Terns, M.P. (2018) The ribonuclease activity of Csm6 is required for anti-plasmid immunity by Type III-A CRISPR-Cas systems. RNA Biol. 16, 449-460.

Fu, B.X.H., Smith, J.D., Fuchs, R.T., Mabuchi, M., Curcuru, J., Robb, G.B., Fire, A.Z. (2019) Target-dependent nickase activities of CRISPR-Cas nucleases Cpf1 and Cas9. Nat. Microbiol. 4, 888–897.

Furuhata, Y., Sakai, A., Murakami, T., Morikawa, M., Nakamura, C., Yoshizumi, T., Fujikura, U., Nishida, K., Kato, Y. (2019) A method using electroporation for the protein delivery of Cre recombinase into cultured *Arabidopsis* cells with an intact cell wall. Sci. Rep. 9, 2163 (doi.org/10.1038/s41598-018-38119-9).

Gaj, T., Guo, J., Kato, Y., Sirk, S.J., Barbas, C.F. (2012) Targeted gene knockout by direct delivery of zinc-finger nuclease proteins. Nat. Meth. 9, 805-807.

Gallego-Bartolomé, J., Gardiner, J., Lu, W., Papikian, A., Ghoshal, B., Kuo, H.Y., Zhao, J. M.-C., Segal, D.J., Jacobsen, S.E. (2018) Targeted DNA demethylation of the *Arabidopsis* genome using the human TET1 catalytic domain. Proc. Natl. Acad. Sci. U.S.A. 115, E2125-E2134 (https://doi.org/10.1073/pnas.1716945115).

Gangopadhyay, S.A., Cox, K.J., Manna, D., Lim, D., Maji, B., Zhou, Q., Choudhary, A. (2019) Precision control of CRISPR-Cas9 using small molecules and light. Biochem 58, 234-244.

Gantz, V.M., Akbari, O.S. (2018) Gene editing technologies and applications for insects. Curr. Opin. Insect. Sci. 28, 66-72.

Gao, F., Shen, X.Z., Jiang, F., Wu, Y., Han, C. (2016) DNA-guided genome editing using the *Natronobacterium gregoryi* Argonaute. Nat. Biotechnol. 34, 768–773.

Gao, H., Smith, J., Yang, M., Jones, S., Djukanovic, V., Nicholson, M.G., West, A., Bidney, D., Falco, S.C., Jantz, D., Lyznik, L.A. (2010) Heritable targeted mutagenesis in maize using a designed endonuclease. Plant J. 61, 176-187.

Gapinske, M., Luu, A., Winter, J., Woods, W.S., Kostan, K.A., Shiva, N., Song, J.S., Perez-Pinera, p. (2018) CRISPR-SKIP: Programmable gene splicing with single base editors. Genome Biol. 19, 107.

Garneau, J.E., Dupuis, M.-E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A.H., Moineau, S. (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature 468, 67–71.

Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc. Natl. Acad. Sci. U.S.A. 109, E2579–E2586.

Gaudelli, N.M., Komor, A.C., Rees, H.A., Packer, M.S., Badran, A.H., Bryson, D.I., Liu, D.R. (2017) Programmable base editing of AT to GC in genomic DNA without DNA cleavage. Nature 551, 464-471.

Gehrke, J.M., Cervantes, O., Clement, M.K., Wu, Y., Zeng, J., Bauer, D.E., Pinello, L., Joung, J.K. (2018) An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. Nat. Biotechnol. 36, 977-982.

Gersbach, C.A., Gaj, T., Gordley, R.M., Mercer, A.C., Barbas, C.F. (2011) Targeted plasmid integration into the human genome by an engineered zinc-finger recombinase. Nucl. Ac. Res. 39, 7868-7878.

Ghezraoui, H., Piganeau, M., Renouf, B., Renaud, J.B., Sallmyr, A., Ruis, B., Oh, S., Tomkinson, A.E., Hendrickson,

E.A., Giovannangeli, C., Jasin, M., Brunet, E. (2014) Chromosomal translocations in human cells are generated by canonical nonhomologous end-joining. Mol. Cell 55, 829-842.

Gidoni, D., Srivastava, V., Carmi, N. (2008) Site-specific excisional recombination strategies for elimination of undesirable transgenes from crop plants. *In vitro* Cell. Dev. Biol. Plant 44, 457-467.

Gilbert, L.A., Horlbeck, M.A., Adamson, ..., Weissman, J.S. (2014) Genome-scale CRISPR-mediated control of gene repression and activation. Cell 159, 647–661.

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

Golicz, A.A., Bayer, P.E., Bhalla, P.L., Batley, J., Edwards, D. (2020) Pangenomics comes of age: From bacteria to plant and animal applications. Trends Genet. 36, 132-145.

Gomaa, A.A., Klumpe, H.E., Luo, M.L., Selle, K., Barrangou, R., Beisel, C.L. (2014) Programmable removal of bacterial strains by use of genome-targeting CRISPR-Cas systems. mBio 5, e00928-13 (doi: 10.1128/mBio.00928-13).

Gootenberg, J.S., Abudayyeh, O.O., Lee, J.W., ..., Zhang, F. (2017) Nucleic acid detection with CRISPR-Cas13a/C2c2. Science 356, 438-442.

Gootenberg, J.S., Abudayyeh, O.O., Kellner, M.J., Joung, J., Collins, J.J., Zhang, F. (2018) Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6. Science 360, 439-444.

Gouble, A., Smith, J., Bruneau, S., Perez, C., Guyot, V., Cabaniols, J-P., Leduc, S., Fiette, L., Avé, P., Micheau, B., Duchateau, P., Pâques, F. (2006) Efficient *in toto* targeted recombination in mouse liver by meganuclease-induced double-strand break. J. Gene Med. 8, 616-622.

Grohmann, L., Keilwagen, J., Duensing, N., Dagand, E., Hartung, F., Wilhelm, R., Bendiek, J., Sprink, T. (2019) Detection and identification of genome editing in plants: Challenges and opportunities. Front. Plant Sci. 10, 236 (doi: 10.3389/fpls.2019.00236).

Grünewald, J., Zhou, R., Garcia, S.P., Iyer, S., Lareau, C.A., Aryee, M.J., Joung, J.K. (2019) Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. Nature 569, 433-437.

Grünewald, J., Zhou, R., Lareau, C.A., Garcia, S.P., Iyer, S., Miller, B.R., Langner, L.M., Hsu, J.Y., Aryee, M.J., Joung, J.K. (2020) A dual-deaminase CRISPR base editor enables concurrent adenine and cytosine editing. Nat. Biotechnol. 38, 861-864.

Gu, F., Lin, Y., Wang, Z., Wu, X., Ye, Z., Wang, Y., Lan, H. (2020) Biological roles of LSD1 beyond its demethylase activity. Cell. Mol. Life Sci. 77, 3341-3350.

Gupta, V., Sengupta, M., Prakash, J., Tripathy, B.C. (2017) Transgenic animals and plants. In: Basic and applied aspects of biotechnology, pp. 103-123. Springer, Singapore.

Hajiahmadi, Z., Movahedi, A., Wei, H., Li, D., Orooji, Y., Ruan, H., Zhuge, Q. (2019) Strategies to increase on-target and reduce off-target effects of the CRISPR/Cas9 system in plants. Int. J. Mol. Sci. 20, E3719.

Hale, C.R., Majumdar, S., Elmore, J., Pfister, N., Compton, M., Olson, S., Resch, A.M., Glover, C.V.C. 3rd, Graveley, B.R., Terns, R.M., Erns, M.P. (2012) Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. Mol. Cell 45, 292–302.

Hale, C.R., Zhao, P., Olson, S., Duff, M.O., Graveley, B.R., Wells, L., Terns, R.M., Terns, M.P. (2009) RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. Cell 139, 945–956.

Halpin, C. (2005) Gene stacking in transgenic plants – the challenge for 21st century plant biotechnology. Plant Biotechnol. J. 3, 141-155.

Han, J., Huang, J. (2020) DNA double-strand break repair pathway choice: The fork in the road. Genome Instab. Dis. 1, 10-19.

Han, W., Li, Y., Deng, L., Feng, M. Peng, W., Hallstrom, S., Zhang, J., Peng, N., Liang, Y.X., White, M.F., She, Q. (2017) A type III-B CRISPR-Cas effector complex mediating massive target DNA destruction. Nucl. Ac. Res. 28, 1983-1993.

Harrington, L.B., Burstein, D., Chen, J.S., Paez-Espino, D, Ma, E., Witte, I.P., Cofsky, J.C., Kyrpides, N.C., Banfield, J.F., Doudna, J.A. (2018) Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. Science 362, 839-842.

Hassan, N., Sasano, Y., Kimura, S., Easmin, F., Ekino, K., Taguchi, H., Harashima, S. (2020) CRISPR-PCDup: A

novel approach for simultaneous segmental chromosomal duplication in *Saccharomyces cerevisiae*. AMB Expr. 10, 27.

Haun, W., Coffman, A., Clasen, B.M., Demorest, Z.L., Lowy, A., Ray, E., Retterath, A., Stoddard, T., Juillerat, A., Cedrone, F., Mathis, L., Voytas, D.F., Zhang, F. (2014) Improved soybean oil quality by targeted mutagenesis of the fatty acid desaturase 2 gene family. Plant Biotechnol. J. 12, 934-940.

Hauschild, J., Petersen, B., Santiago, Y., Queisser, A.L., Carnwath, J.W., Lucas-Hahn, A., Zhang, L., Meng, X., Gregory, P.D., Schwinzer, R., Cost, G.J., Niemann, H. (2011) Efficient generation of a biallelic knockout in pigs using zinc-finger nucleases. Proc. Natl. Acad. Sci. U.S.A. 108, 12013-12017.

He, M.D., Zhang, F.H., Wang, H.L., Wang, H.P., Zhu, Z.Y., Sun, Y.H. (2015) Efficient ligase 3-dependent microhomology-mediated end joining repair of DNA double-strand breaks in zebrafish embryos. Mutation Res. 780, 86-96.

Hedges, D.J., Deininger, P.L. (2007) Inviting instability: Transposable elements, double-strand breaks, and the maintenance of genome integrity. Mutat. Res. 616, 46–59.

Heinz, S., Romanoski, C., Benner, C., Glass, C.K. (2015) The selection and function of cell type-specific enhancers. Nat. Rev. Mol. Cell Biol. 16, 144–154.

Helin K., Dhanak D. (2013) Chromatin proteins and modifications as drug targets. Nature 502, 480–488.

Heller, R.C., Marians, K.J. (2006) Replication fork reactivation downstream of a blocked nascent leading strand. Nature 439, 557–562.

Henry, I.M., Nagalakshmi, U., Lieberman, M.C., Ngo, K.J., Krasileva, K.V., Vasquez-Gross, H., Akhunova, A., Akhunov, E., Dubcovsky, J., Tai, T.H., Comai, L. (2014) Efficient genome-wide detection and cataloging of EMS-induced mutations using exome capture and next-generation sequencing. Plant Cell 26, 1382-1397.

Her, J., Bunting, S.F. (2018) How cells ensure correct repair of DNA double-strand breaks. J. Biol. Chem. 293, 10502-10511.

Hess, G.T., Fresard, L., Han, K., Lee, C.H., Li, A., Cimprich, K.A., Montgomery, S.B., Bassik, M.C. (2016) Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. Nat. Meth. 3, 1036-1042.

Hew, B.E., Sato, R., Mauro, D., Stoytchev, I., Owens, J.B. (2019) RNA-guided piggyBac transposition in human cells. Synth. Biol. (Oxf). 4, ysz018 (doi: 10.1093/synbio/ysz018).

Hickey, L.T, Hafeez, A.N., Robinson, H., Jackson, S.A., Leal-Bertioli, S.C.M., Tester, M., Gao, C., Godwin, I.D., Hayes, B.J., Wulff, B.B.H. (2019) Breeding crops to feed 10 billion. Nat. Biotechnol. 37,744-754.

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

Hilton, I.B., Gersbach, C.A. (2015) Enabling functional genomics with genome engineering. Genome Res. 25, 1442–1455.

Hirano, H., Gootenberg, J.S., Horii, T., Abudayyeh, O.O., Kimura, M., Hsu, P.D., Nakane, T., Ishitani, R., Hatada, I., Zhang, F., Nishimasu, H., Nureki, O. (2016) Structure and engineering of *Francisella novicida* Cas9. Cell 164, 950-961.

Holkers, M., Maggio, I., Liu, J., Janssen, J.M., Miselli, F., Mussolino, C., Recchia, A., Cathomen, T., Gonçalves, M.A. (2013) Differential integrity of TALE nuclease genes following adenoviral and lentiviral vector gene transfer into human cells. Nucl. Ac. Res. 41, 5e63.

Holtzmann, L., Gersbach, C.A. (2018) Editing the epigenome: Reshaping the genomic landscape. Annu. Rev. Genom. Hum. Genet. 19, 43–71.

Howorth, M. (1960) Atomic gardening for the layman. St. Leonard-on-sea, Great Britain: King Brosand Potts LTD.

Hoy, S.M. (2018) Patisiran: First global approval. Drugs 78, 1625-1631.

Hu, J.H., Miller, S.M., Geurts, M.H., Tang, W., Chen, L., Sun, N., Zeina CM, Gao X, Rees HA, Lin Z, Liu DR. (2018) Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. Nature 556, 57-63.

Hu, Q., Lu, H., Wang, H., Li, S., Truong, L., Li, J., Liu, S., Xiang, R., Wu, X. (2019) Break-induced replication plays a prominent role in long-range repeat-mediated deletion. EMBO J. 38, e101751.

Hu, S., Fu, J., Huang, F., Ding, X., Stewart, A.F., Xia, L., Zhang, Y. (2014) Genome engineering of *Agrobacterium*

tumefaciens using the lambda Red recombination system. Appl. Microbiol. Biotechnol. 98, 2165-2172.

Hua, K., Tao, X., Zhu, J.K. (2018) Expanding the base editing scope in rice by using Cas9 variants. Plant Biotechnol. J. 17, 499-504.

Hua, Y., Krainer, A.R. (2012) Antisense-mediated exon inclusion. Meth. Mol. Biol. 867, 307-323.

Huang, T.P., Zhao, K.T., Miller, S.M., Gaudelli, N.M., Oakes, B.L., Fellmann, C., Savage, D.F., Liu, D.R. (2019b) Circularly permuted and PAM-modified Cas9 variants broaden the targeting scope of base editor. Nat. Biotechnol. 37, 626-631.

Huang, Y., Porter, A., Zhang,Y., Barrangou, R. (2019a) Collaborative networks in gene editing. Nat. Biotechnol. 37, 1107-1109.

Hudecek, M., Izsvák, Z., Johnen, S., Renner, M., Thumann, G., Ivics, Z. (2017) Going non-viral: the Sleeping Beauty transposon system breaks on through to the clinical side. Crit. Rev. Biochem. Mol. Biol. 52, 355-380.

Huen, M.S.Y., Li, X.-t., Lu, L.-Y., Watt, R.M., Liu, D.-P., Huang, J.-D. (2006) The involvement of replication in single stranded oligonucleotide-mediated gene repair. Nucl. Ac. Res. 34, 6183-6194.

Hwang, S., Maxwell, K.L. (2019) Meet the anti-CRISPRs: Widespread protein inhibitors of CRISPR-Cas systems. CRISPR J. 2, 23-30.

Hyodo, T., Rahman, M.L., Karnan, S., Ito, T., Toyoda, T., Ota, A., Wahiduzzaman, M., Tsuzuki, S., Okada, Y., Hosokawa, Y., Konishi, H. (2020) Tandem paired nicking promotes precise genome editing with scarce interference by p53. Cell Rep. 30, 1195-1207.

Ichikawa, H.T., Cooper, J.C., Lo, L., Potter, J., Terns, R.M., Terns, M.P. (2017) Programmable type III-A CRISPR-Cas DNA targeting modules. PLoS One 12, e0176221.

Imelfort, M., Edwards, D. (2009) *De novo* sequencing of plant genomes using second-generation technologies. Brief. Bioinform. 10, 609–618.

International Wheat Genome Sequencing Consortium (2018) Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science 361, 661 (DOI: 10.1126/science.aar7191).

Irshad, A., Guo, H., Zhang, S., Liu, L. (2020) TILLING in cereal crops for allele expansion and mutation detection by using modern sequencing technologies. Agron. 10, 405 (doi:10.3390/agronomy10030405).

Jiang, F., Doudna, J.A. (2017) CRISPR–Cas9 structures and mechanisms. Annu. Rev. Biophys. 46, 505-529.

Jiang, S., Shen, Q.W. (2019) Principles of gene editing techniques and applications in animal husbandry. 3 Biotech. 9, 28.

Jiang, W., Bikard, D., Cox, D., Zhang, F., Marraffini, L.A. (2013) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat. Biotechnol. 31, 233-239.

Jiang, W., Samai, P., Marraffini, L.A. (2016) Degradation of phage transcripts by CRISPR-associated RNases enables type III CRISPR-Cas immunity. Cell 164, 710–721.

Jiao, Y., Peluso, P., Shi, J., ..., Ware, D. (2017) Improved maize reference genome with single-molecule technologies. Nature 546, 524-527.

Jin, S., Zong, Y., Gao, Q., Zhu, Z., Wang, Y., Qin, P., Liang, C., Wang, D., Qiu, J.L., Zhang, F., Gao, C. (2019) Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. Science 364, 292-295.

Jinek, M. (2019) Uncut but primed for change. CRISPR J. 2, 352-354.

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

Jones, P.A. (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat. Rev. Genet., 13, 484–492.

Juillerat, A., Dubois, G, Valton, J., Thomas, S., Stella, S., Maréchal, A., Langevin, S., Benomari, N., Bertonati, C., Silva, G.H., Daboussi, F., Epinat, J.C., Montoya, G., Duclert, A., Duchateau, P. (2014) Comprehensive analysis of the specificity of transcription activator-like effector nucleases. Nucl. Ac. Res. 42, 5390-5402.

Juillerat, A., Gronemeyer, T., Keppler, A., Gendreizig, S., Pick, H., Vogel, H., Johnsson, K. (2003) Directed evolution of O6-alkylguanine-DNA alkyltransferase for efficient labeling of fusion proteins with small molecules *in vivo*. Chem. Biol. 10, 313-317.

Jung, Y.-J., Nogoy, F.M., Lee, S.-K., Cho, Y.-G., Kang, K.-K. (2018) Application of ZFN for site directed mutagenesis of rice SSIVa gene. Biotechnol. Bioproc. Eng. 23, 108-115.

Jupe, F., Rivkin, A.C., Michael, T.P., Zander, M., Motley, S.T., Sandoval, J.P., Slotkin, R.K., Chen, H., Castanon, R., Nery, J.R., Ecker, J.R. (2019) The complex architecture and epigenomic impact of plant T-DNA insertions. PLoS Genet. 15, e1007819 (doi: 10.1371/journal.pgen.1007819).

Jurica, M.S., Stoddard, B.L. (1999) Homing endonucleases: Structure, function and evolution. Cell. Mol. Life Sci. 55, 1304-1326.

Jurkowski, T.P., Ravichandran, M., Stepper, P. (2015). Synthetic epigenetics—towards intelligent control of epigenetic states and cell identity. Clin. Epigen. 7, 18.

Kaczmarek, J.C., Kowalski, P.S., Anderson, D.G. (2017) Advances in the delivery of RNA therapeutics: from concept to clinical reality. Genome Med. 9, 60 (doi.org/10.1186/s13073-017-0450-0).

Kar, A., Kuo, D., He, R., Zhou, J., and Wu, J.Y. (2005) Tau alternative splicing and frontotemporal dementia. Alzheimer Dis. Assoc. Disord. 19 (Suppl 1), S29–S36.

Karvelis, T., Bigelyte, G., Young, J.K., Hou, Z., Zedaveinyte, R., Pociute, K., Silanskas, A., Venclovas, Č., Siksnys, V. (2020) PAM recognition by miniature CRISPR-Cas12f triggers programmable double-stranded DNA target cleavage. Nucl. Ac. Res. 48, 5016-5023.

Karvelis, T., Gasiunas, G., Siksny, V. (2017) Harnessing the natural diversity and *in vitro* evolution of Cas9 to expand the genome editing toolbox. Curr. Opin. Microbiol. 37, 88-94.

Katayama, K., Mitsunobu, H., Nishida, K. (2019) Mammalian synthetic biology by CRISPRs engineering and applications. Curr. Op. Chem. Biol. 52, 79–84.

Kazlauskiene, M., Tamulaitis, G., Kostiuk, G., Venclovas, C., Siksnys, V. (2016) Spatiotemporal control of type III-A CRISPR-Cas immunity: Coupling DNA degradation with the target RNA recognition. Mol. Cell 62, 295–306.

Keam, S.J. (2018) Inotersen: first global approval. Drugs 78, 1371-1376.

Kempton, H.R., Qi, L.S. (2019) When genome editing goes off-target. Science 364, 234-236.

Keppler, A., Gendreizig, S., Gronemeyer, T., Pick, H., Vogel, H., Johnsson, K.A. (2002) General method for the covalent labeling of fusion proteins with small molecules *in vivo*. Nat. Biotechnol. 21, 86-89.

Kerschen, P., Planté-Bordeneuve, V. (2016) Current and future treatment approaches in transthyretin familial amyloid polyneuropathy. Curr. Treat. Options Neurol. 18, 53 (doi:10.1007/s11940-016-0436-z).

Kersey, P.J. (2019) Plant genome sequences: past, present, future. Curr. Opin. Plant Biol. 48, 1-8.

Keung, A.J., Bashor, C.J., Kiriakov, S., Collins, J.J., Khalil, A.S. (2014) Using targeted chromatin regulators to engineer combinatorial and spatial transcriptional regulation. Cell 158, 110-120.

Keryer-Bibens, C., Barreau, C., Osborne, B.H. (2008) Tethering of proteins to RNAs by bacteriophage proteins. Biol. Cell 100, 125-138.

Kim, H., Kim, S.T., Ryu, J., Kang, B.C., Kim, J.S., Kim, S.G. (2017a) CRISPR/Cpf1-mediated DNA-free plant genome editing. Nat. Commun. 8, 14406.

Kim, K., Ryu, S.M., Kim, S.T., Baek, G., Kim, D., Lim, K., Chung, E., Kim, S., Kim, J.S. (2017c) Highly efficient RNA-guided base editing in mouse embryos. Nat. Biotechnol. 35, 435-437.

Kim, N., Kim, H.K., Lee, S., Seo, J.H., Choi, J.W., Park, J., Min, S., Yoon, S., Cho, S.-R., Kim, H.H. (2020) Prediction of the sequence-specific cleavage activity of Cas9 variants. Nat. Biotechnol. 38, 1328-1336.

Kim, Y.B., Komor, A.C., Levy, J.M., Packer, M.S., Zhao, K.T., Liu, D.R. (2017b) Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. Nat. Biotechnol. 35, 371-376.

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

Klann, T.S., Black, J.B., Chellappan, M., Safi, A., Song, L., Hilton, I.B., Crawford, G.E., Reddy, T.E., Gersbach, C.A. (2017) CRISPR-Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome. Nat. Biotechnol. 35, 561-568.

Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N., Zheng, Z., Joung, J.K. (2016) High-fidelity CRISPR-Cas9 variants with undetectable genome-wide off-targets. Nature 529, 490-495.

Kleinstiver, B.P., Sousa, A.A., Walton, R.T., Tak, Y.E., Hsu, J.Y., Clement, K., Welch, M.M., Horng, J.E., Malagon-Lopez, J., Scarfò, I., Maus, M.V., Pinello, L., Aryee, M.J., Joung, J.K. (2019) Engineered CRISPR–Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. Nat. Biotechnol. 37, 276-282.

Klompe, S.E., Vo, P.L.H., Halpin-Healy, T.S., Sternberg, S.H. (2019) Transposon-encoded CRISPR–Cas systems direct RNA-guided DNA integration. Nature 571, 219-225.

Koblan, L.W., Doman, J.L., Wilson, C., Levy, J.M., Tay, T., Newby, G.A., Maianti, J.P., Raguram, A., Liu, D.R. (2018) Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. Nat. Biotechnol. 36, 843-846.

Koboldt, D.C., Steinberg, K.M., Larson, D.E., Wilson, R.K., Mardis, E., (2013) The Next-Generation Sequencing revolution and its impact on genomics. Cell 155, 27-38.

Komor, A.C., Badran, A.H., Liu, D.R. (2017) CRISPR-Based technologies for the manipulation of eukaryotic genomes. Cell 168, 20-36.

Komor, A.C., Badran, A.H., Liu, D.R. (2018) Editing the genome without double-stranded DNA breaks. ACS Chem. Biol. 13, 383-388.

Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., Liu, D.R. (2016) Programmable editing of a target base in genomic DNA without double stranded DNA cleavage. Nature 533, 420-424.

Konermann, S., Brigham, M.D., Trevino, A.E., ..., Zhang, F. (2015) Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature 29, 517, 583-588.

Konermann, S., Lotfy, P., Brideau, N.J., Oki, J., Shokhirev, M.N., Hsu, P.D. (2018) Transcriptome engineering with RNA-targeting type VI-D CRISPR effectors. Cell 173, 665–676.

Koonin, E.V., Makarova, K.S. (2019) Origins and evolution of CRISPR-Cas systems. Philos. Trans. R. Soc. Lond. B Biol. Sci. 374, 20180087.

Koonin, E.V., Makarova, K.S., Zhang, F. (2017) Diversity, classification and evolution of CRISPR-Cas systems. Curr. Opin. Microbiol. 37, 67-78.

Kostyrko, K., Neuenschwander, S., Junier, T., Regamey, A., Iseli, C., Schmid-Siegert, E., Bosshard, S., Majocchi, S., Le Fourn, V., Girod, P.A., Xenarios, I., Mermod, N. (2017) MAR-mediated transgene integration into permissive chromatin and increased expression by recombination pathway engineering. Biotechnol. Bioeng. 114, 384-396.

Kovač, A., Miskey, C., Menzel, M., Grueso, E., Gogol-Döring, A., Ivics, Z. (2020) RNA-guided retargeting of Sleeping Beauty transposition in human cells. eLife 9, 53868 (doi:10.7554/eLife.53868).

Kumar, A. (2020) Jump around: transposons in and out of the laboratory. F1000 Research 9 (F1000 Faculty Rev.), 135 (doi.org/10.12688/f1000research.21018.1)

Kumar S., Chinnusamy V., Mohapatra T. (2018) Epigenetics of modified DNA bases: 5-methylcytosine and beyond. Front. Genet. 9, 640.

Kungulowski, G., Jeltsch, A. (2016) Epigenome editing: State of the art, concepts, and perspectives. Trends Genet. 32, 101-113.

Kurreck, J. (2003) Antisense technologies. Improvement through novel chemical modifications. Eur. J. Biochem. 270, 1628–1644.

Kurt, I.C., Zhou, R., Iyer, S., Garcia, S.P., Miller, B.R., Langner, L.M., Grünewald, J., Joung, J.K. (2020) CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. Nat. Biotechnol. 39, 41-46.

Kuscu, C., Parlak, M., Tufan, T., Yang, J., Szlachta, K., Wei, X., Mammadov, R., Adli, M. (2017) CRISPR-STOP: Gene silencing through base-editing-induced nonsense mutations. Nat. Methods 14, 710-712.

Kyriakidou, M., Tai, H.H., Anglin, N.L., Ellis, D., Strömvik, M.V. (2018) Current strategies of polyploid plant genome sequence assembly. Front. Plant Sci. 9, 1660 (doi.org/10.3389/fpls.2018.01660).

Lacchini, E., Kiegle, E., Castellani, M., Adam, H., Jouannic, S., Gregis, V., Kater, M.M. (2020) CRISPR-mediated accelerated domestication of African rice landraces. PLoS One 15, e0229782.

Landy, A. (2015) The λ Integrase Site-specific Recombination Pathway. Microbiol. Spectr. 3, MDNA3–0051-2014 (doi:10.1128/microbiolspec.MDNA3-0051-2014).

Lee, J., Rho, J., Devkota, S., Sung, Y.H., Woong, H.L. (2014) Developing genetically engineered mouse models

using engineered nucleases: Current status, challenges, and the way forward. Drug Discov. Today: Dis. Mod. 20, 13-20.

Lee, K., Uh, K., Farrell, K. (2020) Current progress of genome editing in livestock. Theriogenol. 150, 229-235.

Leenay, R.T., Beise, C.L. (2017) Deciphering, communicating, and engineering the CRISPR PAM. J. Mol. Biol. 429, 177-191.

Li, C., Zhang, R., Meng, X., Chen, S., Zong, Y., Lu, C., Qiu, J.-L., Chen, Y.-H., Li, J., Gao, C. (2020) Targeted, random mutagenesis of plant genes with dual cytosine and adenine base editors. Nat. Biotechnol. 38, 875–882.

Li, C., Zong, Y., Wang, Y., Jin, S., Zhang, D., Song, Q., Zhang, R., Gao, C. (2018b) Expanded base editing in rice and wheat using a Cas9-adenosine deaminase fusion. Genome Biol. 19, 59.

Li, G., Chern, M., Jain, R., Martin, J.A., Schackwitz, W.S., Jiang, L., Vega-Sánchez, M.E., Lipzen, A.M., Barry, K.W., Schmutz, J., Ronald, P.C. (2016a) Genome-wide sequencing of 41 rice (*Oryza sativa* L.) mutated lines reveals diverse mutations induced by fast-neutron irradiation. Mol. Plant. 9, 1078-1081.

Li, J., Hong, S., Chen, W., Zuo, E., Yang, H. (2019a) Advances in detecting and reducing off-target effects generated by CRISPR-mediated genome editing. J. Genet. Genom. 46, 513-521.

Li, J., Wang, X., Wang, X., Wang, Y. (2015) Embryo rescue technique and its applications for seedless breeding in grape. Plant Cell Tiss. Organ Cult. 120, 861–880.

Li, M., Chen, L., Tian, S., Lin, Y., Tang, Q., Zhou, X., Li, D., Yeung, C., Che, T., Jin, L., Fu, Y., Ma, J., Wang, X., Jiang, A., Lan, J., Pan, Q., Liu, Y., Luo, Z., Guo, Z., Liu, H., … Li, X. (2017) Comprehensive variation discovery and recovery of missing sequence in the pig genome using multiple *de novo* assemblies. Genome Res. 27, 865-874.

Li, S., Li, J., He, Y., Xu, M., Zhang, J., Du, W., Zhao, Y., Xia, L. (2019b) Precise gene replacement in rice by RNA transcript-templated homologous recombination. Nat. Biotechnol. 37, 445-450.

Li, T., Liu, B., Chen, C.Y., Yang, B. (2016b) TALEN-mediated homologous recombination produces site-directed DNA base change and herbicide-resistant rice. J. Genet. Genom. 43, 297-305.

Li, T., Liu, B., Spalding, M.H., Weeks, D.P., Yang, B. (2012) High-efficiency TALEN-based gene editing produces disease-resistant rice. Nat. Biotechnol. 30, 390-392.

Li, X., Bai, Y., Cheng, X., Kalds, P.G.T., Sun, B., Wu, Y., Lv, H., Xu, K., Zhang, Z. (2018a) Efficient SSA-mediated precise genome editing using CRISPR/Cas9. FEBS J. 285, 3362-3375.

Li, X., Wang, Y., Liu, Y., Yang, B., Wang, X., Wei, J.,Lu, Z., Zhang, Y., Wu, J., Huang, X., Yang, L., Chen, J. (2018c) Base editing with a Cpf1–cytidine deaminase fusion. Nat. Biotechnol. 36, 324-327.

Li, Y., Li, S., Wang, J., Liu, G. (2019c) CRISPR/Cas systems towards next-generation biosensing. Trends Biotechnol. 37, 730-743.

Li, Y.H., Zhou, G., Ma, J., …, Qiu, L.-j. (2014) *De novo* assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. Nat. Biotechnol. 32, 1045-1052.

Li, Z., Pearlman, A.H., Hsieh, P. (2016c) DNA mismatch repair and the DNA damage response. DNA Repair 38, 94-101.

Liang, P., Ding, C., Sun, H., Xie, X., Xu, Y., Zhang, X., Sun, Y., Xiong, Y., Ma, W., Liu, Y., Wang, Y., Fang, J., Liu, D., Songyang, Z., Zhou, C., Huang, J. (2017b) Correction of beta-thalassemia mutant by base editor in human embryos. Protein Cell 8, 811-822.

Liang, X., Potter, J., Kumar, S., Zou, Y., Quintanilla, R., Sridharan, M., Carte, J., Chen, W., Roark, N., Ranganathan, S., Ravinder, N., Chesnut, J.D. (2015) Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. J. Biotechol. 208, 44-53.

Liang, X.H., Hart, C.H., Crooke, S.T. (2013) Transfection of siRNAs can alter miRNA levels and trigger non-specific protein degradation in mammalian cells. Biochim. Biophys. Acta 1829, 455–468.

Liang, Z., Chen, K., Li, T., Zhang, Y., Wang, Y., Zhao, Q., Liu, J., Zhang, H., Liu, C., Ran, Y., Gao, C. (2017a) Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes. Nat. Commun. 8, 14261.

Lieber, M.R. (2010) The mechanism of double-strand DNA break repair by the nonhomologous DNA end joining pathway. Annu. Rev. Biochem. 79, 181-211.

Lillico, S.G., Proudfoot, C., King, T.J., Tan, W., Zhang, L., Mardjuki, R., Paschon, D.E., Rebar, E.J., Urnov, F.D., Mileham, A.J., McLaren, D.G., Whitelaw, C.B. (2016) Mammalian interspecies substitution of immune

modulatory alleles by genome editing. Sci. Rep. 6, 21645.

Lima, W.F., Prakash, T.P., Murray, H.M., ..., Crooke, S.T. (2012) Single-stranded siRNAs activate RNAi in animals. Cell 150, 883–894.

Lin, Y., Cradick, T.J., Brown, M.T., Deshmukh, H., Ranjan, P., Sarode, N., Wile, B.M., Vertino, P.M., Stewart, F.J., Bao, G. (2014) CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. Nucl. Ac. Res. 42, 7473-7485.

Liu, C., Zhang, L., Liu, H., Cheng, K. (2017a) Delivery strategies of the CRISPR-Cas9 gene-editing system for therapeutic applications. J. Contr. Rel. 266, 17–26.

Liu, Q., Segal, D.S., Ghiara, J.B., Barbas III, C.F. (1997) Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. Proc. Natl. Acad. Sci. U.S.A. 94, 5525-5530.

Liu, T., Pan, S., Li, Y., Peng, N., She, Q. (2018c) Type III CRISPR-Cas system: Introduction and its application for genetic manipulations. Curr. Iss. Mol. Biol. 26, 1–14.

Liu, T.Y., Iavarone, A.T., Doudna, J.A. (2017b) RNA and DNA targeting by a reconstituted *Thermus thermophiles* type III-A CRISPR-Cas system. PLoS One 12, e0170552.

Liu, X.S., Wu, H., Ji, X., Stelzer, Y., Wu, X., Czauderna, S., Shu, J., Dadon, D., Young. R.A., Jaenisch, R. (2016) Editing DNA methylation in the mammalian genome. Cell 167, 233–247.

Liu, Z., Chen, M., Chen, S., Deng, J., Song, Y., Lai, L., Li, Z. (2018a) Highly efficient RNA-guided base editing in rabbit. Nat. Commun. 9, 2717.

Liu, Z., Lu, Z., Yang, G., Huang, S., Li, G., Feng, S., Liu, Y., Yu, W., Zhang, Y., Chen, J., Sun, Q., Huang, X. (2018b) Efficient generation of mouse models of human diseases via ABE- and BE-mediated base editing. Nat. Commun. 9, 2338.

Liu, Z.H., Xing, J.F., Chen, W., Luan, M.W., Xie, R., Huang, J., Xie, S.-Q., Xiao, C.-L. (2019) MDR: an integrative DNA N6-methyladenine and N4- methylcytosine modification database for Rosaceae. Hort. Res. 6, 78.

Lomas, D.A., Mahadeva, R. (2002) α 1-antitrypsin polymerization and the serpinopathies: pathobiology and prospects for therapy. J. Clin. Invest. 110, 1585–1590.

Long, C., Li, H., Tiburcy, M., Olson, E.N. (2018) Correction of diverse muscular dystrophy mutations in human engineered heart muscle by single-site genome editing. Sci. Adv. 4, eaap9004 (doi: 10.1126/sciadv.aap9004).

Lu, X., Zhao, B.X., He, C. (2015) TET family proteins: Oxidation activity, interacting molecules, and functions in diseases. Chem. Rev. 115, 2225–2239.

Lu, Y., Zhu, J.K. (2017) Precise editing of a target base in the rice genome using a modified CRISPR/Cas9 system. Mol. Plant 10, 523-525.

Luker, K.E., Smith, M.C.P., Luker, G.D., Gammon, S.T., Piwnica-Worms, H., Piwnica-Worms, D. (2004) Kinetics of regulated protein–protein interactions revealed with firefly luciferase complementation imaging in cells and living animals. Proc. Natl. Acad. Sci. U.S.A. 101, 12288–12293.

Lusser, M., Parisi, C., Plan, D., Rodríguez-Cerezo, E. (2011) New plant breeding techniques. State-of-the-art and prospects for commercial development. Luxembourg, Publ. Off. Eur. Union, 184 pp. (doi:10.2791/54761).

Lusser, M., Parisi, C., Plan, D., Rodríguez-Cerezo, E. (2012) Deployment of new biotechnologies in plant breeding. Nat. Biotechnol. 30, 231–239.

Ma, X., Zhang, Q., Zhu, Q., ..., Liu, Y.G. (2015) A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. Mol. Plant 8, 1274-1284.

Ma, X., Zhang, X., Liu, H., Li, Z. (2020) Highly efficient DNA-free plant genome editing using virally delivered CRISPR–Cas9. Nat. Plants 6, 773–779.

Ma, Y., Yu, L., Zhang, X., Xin, C., Huang, S., Bai, L., Chen, W., Gao, R., Li, J., Pan, S., Qi, X., Huang, X., Zhang, L. (2018) Highly efficient and precise base editing by engineered dCas9-guide tRNA adenosine deaminase in rats. Cell Discov. 4, 39.

Ma, Y., Zhang, J., Yin, W., Zhang, Z., Song, Y., Chang, X. (2016) Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. Nat. Meth. 13, 1029-1035.

Maeder, M., Angstman, J.F., Richardson, M.E., Linder, S.J., Cascio, V.M., Tsai, S.Q., Ho, Q.H., Sander, J.D., Reyon, D., Bernstein, B.E., Costello, J.F., Wilkinson, M.F., Joung, J.K. (2013a) Targeted DNA demethylation and

activation of endogenous genes using programmable TALE-TET1 fusion proteins. Nat. Biotechnol. 31, 1137-1142.

Maeder, M., Linder, S.J., Cascio, V.M., Fu, Y., Ho, Q.H., Joung, J.K. (2013b) CRISPR RNA-guided activation of endogenous human genes. Nat. Meth. 10, 977-979.

Maeder, M.L., Gersbach, C.A. (2016) Genome-editing technologies for gene and cell therapy. Mol. Ther. 24, 430-446.

Malnoy, M., Viola, R., Jung, M.H., Koo, O.J., Kim, S., Kim, J.S., Velasco, R., Nagamangala Kanchiswamy, C. (2016) DNA-free genetically edited grapevine and apple protoplast using CRISPR/Cas9 ribonucleoproteins. Front. Plant Sci. 7, 1904.

Manghwar, H., Li, B., Ding, X., Hussain, A., Lindsey, K., Zhang, X., Jin, S. (2020) CRISPR/Cas systems in genome editing: Methodologies and tools for sgRNA design, off-target evaluation, and strategies to mitigate off-target effects. Adv. Sci. (Weinh) 7, 1902312.

Manova, V., Gruszka, D. (2015) DNA damage and repair in plants - from models to crops. Front. Plant Sci. 6, 885.

Maragathavally KJ, Kaminski JM, Coates CJ. (2006) Chimeric Mos1 and piggyback transposases result in site-directed integration. FASEB J. 20, 1880-1882.

Marton, I., Zuker, A., Shklarman, E., Zeevi, V., Tovkach, A., Roffe, S., Ovadis, M., Tzfira, T., Vainstein, A. (2010) Nontransgenic genome modification in plant cells. Plant Physiol. 154, 1079–1087.

Maslowska, K.H, Makiela-Dzbenska, K., Fijalkowska, I.J. (2019) The SOS system: A complex and tightly regulated response to DNA damage. Environ. Mol. Mutagen 60, 368-384.

McKenna, A., Findlay, G.M., Gagnon, J.A., Horwitz, M.S., Schier, A.F., Shendure, J. (2016) Whole-organism lineage tracing by combinatorial and cumulative genome editing. Science 353, aaf7907 (doi:10.1126/science.aaf7907).

Mehta, A. Haber, J.E. (2014) Sources of DNA double-strand breaks and models of recombinational DNA repair. Cold Spring Harb. Perspect. Biol. 6, a016428.

Mejías-Navarro, F., Rodríguez-Real, G., Ramón, J., Camarillo, R., Huertas, P. (2020) ALC1/eIF4A1-mediated regulation of CtIP mRNA stability controls DNA end resection. PLoS Genet. 16, e1008787.

Meinke, G., Bohm, A., Hauber, J., Pisabarro, M.T., Buchholz, F. (2016) Cre recombinase and other tyrosine recombinases. Chem. Rev. 116, 12785-12820.

Mendenhall, E., Williamson, K., Reyon, D., Zou, J.Y., Ram, O., Joung, J.K., Bernstein, B.E. (2013) Locus-specific editing of histone modifications at endogenous enhancers. Nat. Biotechnol. 31, 1133–1136.

Mercer, A.C., Gaj, T., Fuller, R.P, Barbas, C.F. (2012) Chimeric TALE recombinases with programmable DNA sequence specificity. Nucl. Ac. Res. 40, 11163-11172.

Mercuri, E., Darras, B.T., Chiriboga, C.A., ... Finkel, R.S. (2018) Nusinersen versus sham control in later-onset spinal muscular atrophy. New Engl. J. Med. 378, 625–635.

Merkle, T., Merz, S., Reautschnig, P., Blaha, A., Li, Q., Vogel, P., Wettengel, J., Li, J.B., Stafforst, T. (2019) Precise RNA editing by recruiting endogenous ADARs with antisense oligonucleotides. Nat. Biotechnol. 37, 133-138.

Metzger, M.J., McConnell-Smith, A., Stoddard, B.L., Miller, A.D. (2011) Single-strand nicks induce homologous recombination with less toxicity than double-strand breaks using an AAV vector template. Nucl. Ac. Res. 39, 926-935.

Miglani, G.S., Kaur, A., Kaur, L. (2020) Plant gene expression control using genome- and epigenome-editing technologies. J. Crop Improv. 34, 1-63.

Miller, J.C., Holmes, M.C., J. Wang, J., Guschin, D.Y., Lee, Y.L., Rupniewski, I., Beausejour, C.M., Waite, A.J., Wang, N.S., Kim, K.A., Gregory, P.D., Pabo, C.O., Rebar, E.J. (2007) An improved zinc-finger nuclease architecture for highly specific genome editing. Nat. Biotechnol. 25, 778-785.

Minorikawa, S., Nakayama, M. (2011) Recombinase-mediated cassette exchange (RMCE) and BAC engineering via VCre/VloxP and SCre/SloxP systems. Biotechniques 50, 235-246.

Mishra, R., Joshi, R.K., Zhao, K. (2020) Base editing in crops: current advances, limitations and future implications. Plant Biotechnol. J. 18, 20–31.

Miura, T., Yamana, Y., Usui, T., Ogawa, H.I., Yamamoto, M.T., Kusano, K. (2012) Homologous recombination via

synthesis-dependent strand annealing in yeast requires the Irc20 and Srs2 DNA helicases. Genetics 191, 65-78.

Miyaoka, Y., Berman, J.R., Cooper, S.B., Mayerl, S.J., Chan, A.H., Zhang, B., Karlin-Neumann, G.A., Conklin, B.R. (2016) Systematic quantification of HDR and NHEJ reveals effects of locus, nuclease, and cell type on genome editing. Sci. Rep. 6, 23549.

Mizuno, S., Takami, K., Daitoku, Y., Tanimoto, Y., Huong Dinh, T.T., Mizuno-Lijima, S., Hasegawa, Y., Takahasshi, S., Sugiyama, F., Yagami, K.I. (2015) Peri-implantation lethality in mice carrying megabase-scale deletion on 5qc3.3 is caused by Exoc1 null mutation. Sci. Rep. 5, 13632.

Moghaddassi, S.M., Eyestone, W., Bishop, C.E. (2014) TALEN-Mediated Modification of the bovine genome for large-scale production of human serum albumin. PLoS One 9, e89631.

Mojica, F.J.M., Diez-Villasenor, C., Garcia-Martinez, J., Almendros, C. (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiol. 155, 733–740.

Mojica, F.J.M., Montoliu, L. (2016) On the origin of CRISPR-Cas technology: From prokaryotes to mammals. Trends Microbiol. 24, 811-820.

Mok, B.Y., de Moraes, M.H., Zeng, J., Bosch, D.E., Kotrys, A.V., Raguram, A., Hsu, F., Radey, M.C., Peterson, S.B., Mootha, V.K., Mougous, J.D., Liu, D.R. (2020) A bacterial cytidine deaminase toxin enables CRISPR-free mitochondrial base editing. Nature 583, 631–637.

Montiel-Gonzalez, M.F., Diaz Quiroz, J.F., Rosenthal, J.J.C. (2019) Current strategies for site-directed RNA editing using ADARs. Methods 156, 16-24.

Montiel-Gonzalez, M.F., Vallecillo-Viejo, I.C., Yudowski, G.A., Rosenthal. J.J.C. (2013) Correction of mutations within the cystic fibrosis transmembrane conductance regulator by site-directed RNA editing. Proc. Natl. Acad. Sci. U.S.A. 110, 18285-18290.

Montiel-Gonzalez, M.F., Vallecillo-Viejo, I.C., Rosenthal, J.J.C. (2016) An efficient system for selectively altering genetic information within mRNAs. Nucl. Ac. Res. 44, e157 (doi: 10.1093/nar/gkw738).

Montoliu, L., Whitelaw, C.B.A. (2018) Unexpected mutations were expected and unrelated to CRISPR-Cas9 activity. Transg. Res. 27, 315-319.

Morgan, S.L., Mariano, N.C., Bermudez, A., ..., Wang, K.C. (2017) Manipulation of nuclear architecture through CRISPR-mediated chromosomal looping. Nat. Commun. 8, 15993 (doi: 10.1038/ncomms15993).

Morio, F., Lombardi, L., Butler, G. (2020) The CRISPR toolbox in medical mycology: State of the art and perspectives. PLoS Pathog. 16, e1008201.

Morita, S., Noguchi, H., Horii, T., Nakabayashi, K., Kimura, M., Okamura, K., Sakai, A., Nakashima, H., Hata, H., Nakashima, K., Hatada, I. (2016) DNA demethylation *in vivo* using dCas9–peptide repeat and scFv–TET1 catalytic domain fusions. Nat. Biotechnol. 34, 1060-1065.

Mou, H., Smith, J.L., Peng, L., ..., Xue, W. (2017) CRISPR/Cas9-mediated genome editing induces exon skipping by alternative splicing or exon deletion. Genome Biol. 18, 108 (doi.org/10.1186/s13059-017-1237-8).

Muñoz-López, M., García-Pérez, J.L. (2010) DNA Transposons: nature and applications in genomics. Curr. Genom. 11, 115–128.

Myhrvold, C., Freije, C.A., Gootenberg, J.S., ..., Sabeti, P.C. (2018) Field-deployable viral diagnostics using CRISPR-Cas13. Science 360, 444-448.

Najera, V.A., Twyman, R.M., Christou, P., Zhu, C. (2019) Applications of multiplex genome editing in higher plants. Curr. Opin. Biotechnol. 59, 93-102.

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

Naldini, L. (2015) Gene therapy returns to centre stage. Nature 526, 351-360.

Nambiar, T.S., Billon, P., Diedenhofen, G., Hayward, S.B., Taglialatela, A., Cai, K., Huang, J.-W., Leuzzi, G., Cuella-Martin, R., Palacios, A., Gupta, A., Egli, D., Ciccia , A. (2019) Stimulation of CRISPR-mediated homology-directed repair by an engineered RAD18 variant. Nat. Commun. 10, 3395 (doi.org/10.1038/s41467-019-11105-z).

Nelles, D.A., Fang, M.Y., Aigner, S., Yeo, G.W. (2015) Applications of Cas9 as an RNA-programmed RNA-binding protein. Bioessays 37, 732-739.

Nelles, D.A., Fang, M.Y., O'Connell, M.R., Xu, J.L., Markmiller, S.J., Doudna, J.A., Yeo, G.W. (2016) Programmable RNA tracking in live cells with CRISPR/Cas9. Cell 165, 488-496.

Nishikura, K. (2010) Functions and regulation of RNA editing by ADAR deaminases. Annu. Rev. Biochem. 79, 321-349.

Nishikura, K. (2016) A-to-I editing of coding and non-coding RNAs by ADARs. Nat. Rev. Mol. Cell Biol. 17, 83-96.

Nishimasu, H., Shi, X., Ishiguro, S., ..., Nureki, O. (2018) Engineered CRISPR-Cas9 nuclease with expanded targeting space. Science 361, 1259-1262.

Niu, D., Wei, H.-J., Lin, ..., Yang, L. (2017) Inactivation of porcine endogenous retrovirus in pigs using CRISPR-Cas9. Science 357, 1303-1307.

Nomura, W. (2018) Development of toolboxes for precision genome/epigenome editing and imaging of epigenetics. Chem. Rec. 18, 1717–1726.

Nomura, W., Barbas, C.F.3rd (2007) *In vivo* site-specific DNA methylation with a designed sequence-enabled DNA methylase. J. Am. Chem. Soc. 129, 8676–8677.

O'Connell, M.R., Oakes, B.L., Sternberg, S.H., East-Seletsky, A., Kaplan, M., Doudna, J.A. (2014) Programmable RNA recognition and cleavage by CRISPR/Cas9. Nature 516, 263-266.

O'Geen, H., Ren, C., Nicolet, C.M., Perez, A.A., Halmai, J., Le, V.M., Mackay, J.P., Farnham, P.J., Segal, D.J. (2017) dCas9-based epigenome editing suggests acquisition of histone methylation is not sufficient for target gene repression. Nucl. Ac. Res. 45, 9901–9916.

Oldenburg, K.R., Vo, K.T., Michaelis, S., Paddon, C. (1997) Recombination-mediated PCR-directed plasmid construction *in vivo* in yeast. Nucl. Ac. Res. 25, 451-452.

Olina, A.V., Kulbachinskiy, A.V., Aravin, A.A., Esyunina, D.M. (2018) Argonaute proteins and mechanisms of RNA interference in eukaryotes and prokaryotes. Biochem. Moscow 83, 483–497.

Ophinni, Y., Palatini, U., Hayashi, Y., Parrish, N.F. (2019) piRNA-guided CRISPR-like immunity in eukaryotes. Trends Immunol. 40, 998-1010.

Orel, N., Kyryk, A., Puchta, H. (2003) Different pathways of homologous recombination are used for the repair of double-strand breaks within tandemly arranged sequences in the plant genome. Plant J. 35, 604-612.

Osawa, T., Inanaga, H., Sato, C., Numata, T. (2015) Crystal structure of the CRISPR-Cas RNA silencing Cmr complex bound to a target analog. Mol. Cell 58, 418–430.

O'Shea, J.J., Schwartz, D.M., Villarino, A.V., Gadina, M., McInnes, J.B., Laurence, A. (2015) The JAK-STAT pathway: impact on human disease and therapeutic intervention. Annu. Rev. Med. 66, 311–328.

Ossowski, S., Schneeberger, K., Lucas-Lledó, J.I., Warthmann, N., Clark, R.M., Shaw, R.G., Weigel, D., Lynch, M. (2010) The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. Science 327, 92-94.

Østerberg, J, W., Olsen, L.I., Edenbrandt, ..., Palmgren, M. (2017) Accelerating the domestication of new crops: Feasibility and approaches. Trends Plant Sci. 22, 373-384.

Owens, J.B., Mauro, D., Stoytchev, I., Bhakta, M.S., Kim, M.-S., Segal, D.J., Moisyadi, S. (2013) Transcription activator like effector (TALE)-directed piggyBac transposition in human cells. Nucl. Ac. Res. 41, 9197-9207.

Pacher, M., Puchta, H. (2017) From classical mutagenesis to nuclease-based breeding – directing natural DNA repair for a natural end-product. Plant J. 90, 819-833.

Pannunzio, N.R., Watanabe, G., Liebe, M.R. (2018) Nonhomologous DNA end-joining for repair of DNA double-strand breaks. J. Biol. Chem. 293, 10512-10523.

Papaioannou, I., Simons, J.P., Owen, J.S. (2012) Oligonucleotide-directed gene-editing technology: Mechanisms and future prospects. Expert Opin. Biol. Ther. 12, 329-342.

Papikian, A., Liu, W., Gallego-Bartolomé, J., Jacobsen, S.E. (2019) Site-specific manipulation of *Arabidopsis* loci using CRISPR-Cas9 SunTag systems. Nat. Commun. 10, 729 (doi.org/10.1038/s41467-019-08736-7).

Pâques, F., Duchateau, P. (2007) Meganucleases and DNA double-strand break-induced recombination: Perspectives for gene therapy. Curr. Gene Ther. 7, 49-66.

Parekh-Olmedo, H., Ferrara, L. Brachman, E., Kmiec, E.B. (2005) Gene therapy progress and prospects: targeted gene repair. Gene Ther. 12, 639-646.

Park, J.J., Dempewolf, E., Zhang, W., Wang, Z.Y. (2017) RNA-guided transcriptional activation via CRISPR/dCas9 mimics overexpression phenotypes in *Arabidopsis*. PLoS One 12, e0179410 (doi:10.1371/journal.pone.0179410).

Parry, M.A.J., Madgwick, P.J., Bayon, C., Tearall, K., Hernandez-Lopez, A., Baudo, M., Rakszegi, M., Hamada, W., Al-Yassin, A., Ouabbou, H., Labhilili, M., Phillips, A.L. (2009) Mutation discovery for crop improvement. J. Exp. Bot. 60, 2817-2825.

Passini, M.A., Bu, J., Richards, A.M., …, Cheng, S.H. (2011) Antisense oligonucleotides delivered to the mouse CNS ameliorate symptoms of severe spinal muscular atrophy. Sci. Transl. Med. 72, 72ra18 (doi: 10.1126/scitranslmed.3001777).

Paszkowski, J., Baur, M., Bogucki, A., Potrykus, I. (1988) Gene targeting in plants. EMBO J. 7, 4021-4026.

Pattanayak, V., Lin, S., Guilinger, J.P., Ma, E., Doudna, J.A., Liu, D.R. (2013) High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. Nat. Biotechnol. 31, 839-843.

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

Paul, B., Montoya, G. (2020) CRISPR-Cas12a: Functional overview and applications. Biomed. J. 43, 8-17.

Paul, C.P., Good, P.D., Winer, I., Engelke, D.R. (2002) Effective expression of small interfering RNA in human cells. Nat. Biotechnol. 20, 505-508.

Pei, W.-D., Zhang, Y., Yin, T.-L., Yang, Y. (2020) Epigenome editing by CRISPR/Cas9 in clinical settings: possibilities and challenges. Brief. Funct. Genom. 19, 215–228.

Peters, J.E., Makarova, K.S., Shmakov, S., Koonin, E.V. (2017) Recruitment of CRISPR-Cas systems by Tn7-like transposons. Proc. Natl. Acad. Sci. U.S.A. 114, E7358-E7366.

Pflueger, C., Tan, D., Swain, T., Nguyen, T., Pflueger, J., Nefzger, C., Polo, J.M., Ford, E., Lister, R. (2018) A modular dCas9-SunTag DNMT3A epigenome editing system overcomes pervasive off-target activity of direct fusion dCas9-DNMT3A constructs. Genome Res. 28, 1193–1206.

Pickar-Oliver, A., Gersbach, C.A. (2019) The next generation of CRISPR–Cas technologies and applications. Nat. Rev. Mol. Cell. Biol. 20, 490-507.

Platt, R.J. (2019) CRISPR tool enables precise genome editing. Nature 576, 48-49.

Pluciennik, A., Dzantiev, L., Iyer, R. R., Constantin, N., Kadyrov, F. A., & Modrich, P. (2010) PCNA function in the activation and strand direction of MutLα endonuclease in mismatch repair. Proc. Natl. Acad. Sci. U.S.A. 107, 16066–16071.

Porter, S.N., Levine, R.M., Pruett-Miller, S.M. (2019) A practical guide to genome editing using targetednuclease technologies. Compr. Physiol. 9, 665-714.

Price, A.A., Sampson, T.R., Ratner, H.K., Grakoui, A., Weiss, D.S. (2015) Cas9-mediated targeting of viral RNA in eukaryotic cells. Proc. Natl. Acad. Sci. U.S.A. 112, 6164-6169.

Puchta, H. (2005) The repair of double-strand breaks in plants: Mechanisms and consequences for genome evolution. J. Exp. Bot. 56, 1-14.

Pyne, M.E., Moo-Young, M., Chung, D.A., Chou, C.P. (2015) Coupling the CRISPR/Cas9 system with lambda Red recombineering enables simplified chromosomal gene replacement in *Escherichia coli*. Appl. Environ. Microbiol. 81, 5103-5114.

Qi, L.L., Rabinowitz, A., Yun, C.C., Yokota, T., Yin, H.F., Alter, J., Jadoon, A., Bou-Gharios, G., Partridge, T. (2005) Systemic delivery of antisense oligoribonucleotide restorers dystrophin expression in body-wide skeletal muscles. Proc. Natl. Acad. Sci. U.S.A. 102, 198–203.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., Lim, W.A. (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell 152, 1173-1183.

Qian, L., Tang, M., Yang, J., Wang, Q., Cai, C., Jiang, S., Li, H., Jiang, K. Gao, P., Ma, D., Chen, Y., An, X., Li, K., Cui, W. (2015) Targeted mutations in myostatin by zinc-finger nucleases result in double-muscled phenotype in Meishan pigs. Sci. Rep. 5, 14435.

Qin, L., Li, J., Wang, Q., Xu, Z., Sun, L., Alariqi, M., Manghwar, H., Wang, G., Li, B., Ding, X., Rui, H., Huang, H., Lu, T., Lindsey, K., Daniell, H., Zhang, X., Jin, S. (2020) High efficient and precise base editing of C•G to T•A in the allotetraploid cotton (*Gossypium hirsutum*) genome using a modified CRISPR/Cas9 system. Plant Biotechnol. J. 18, 45-56.

Rajeevkumar, S., Anunanthini, P., Sathishkumar, R. (2015) Epigenetic silencing in transgenic plants. Front. Plant Sci. 6, 693 (doi:10.3389/fpls.2015.00693).

Ramirez, C.L., Certo, M.T., Mussolino, C., Goodwin, M.J., Cradick, T.J., McCaffrey, A.P., Cathomen, T., Scharenberg, A.M., Joung, J.K. (2012) Engineered zinc finger nickases induce homology-directed repair with reduced mutagenic effects. Nucl. Ac. Res. 40, 5560-5568.

Ran, F.A., Hsu, P.D., Lin, C.-Y., Gootenberg, J.S., Konermann, S., Trevino, A., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y., Zhang, F. (2013) Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell 154, 1380-1389.

Ran, Y., Liang, Z., Gao, C. (2017) Current and future editing reagent delivery systems for plant genome editing. Sci. China Life Sci. 60, 490-505.

Rand, T.A., Petersen, S., Du, F., Wang, X. (2005) Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. Cell 123, 621–629.

Rees, H.A., Wilson, C., Doman, J. L., Liu, D.R. (2019a) Analysis and minimization of cellular RNA editing by DNA adenine base editors. Sci. Adv. 5, eaax5717.

Rees, H.A., Yeh, W.-H., Liu, D.R. (2019b) Development of hRad51-Cas9 nickase fusions that mediate HDR without double-stranded breaks. Nat. Commun. 10, 2212 (doi: 10.1038/s41467-019-09983-4).

Reid, W., O'Brochta, D.A. (2016) Applications of genome editing in insects. Curr. Opin. Insect. Sci. 13, 43-54.

Ren, J., Lee, J., Na, D. (2020) Recent advances in genetic engineering tools based on synthetic biology. J. Microbiol. 58, 1-10.

Richardson, C.D., Kazane, K.R., Feng, S.J., Zelin, E., Bray, N.L., Schäfer, A.J., Floor, S.N., Corn, J.E. (2018) CRISPR–Cas9 genome editing in human cells occurs via the Fanconi anemia pathway. Nat. Genet. 50, 1132-1139.

Rivera-Torres, N., Kmiec, E.B. (2016) Genetic spell-checking: Gene editing using single-stranded DNA oligonucleotides. Plant Biotechnol. J. 14, 463-470.

Roca Paixão, J.F., Gillet, F., Ribeiro, T.P., Bornaud, C., Lourenço-Tessutti, I.T., Noriega, D.D., Paes de Melo, B., de Almeida-Engler, J., Grossi-de-Sa, M.F. (2019) Improved drought stress tolerance in *Arabidopsis* by CRISPR/dCas9 fusion with a histone acetyltransferase. Sci. Rep. 9, 8080 (doi.org/10.1038/s41598-019-44571-y).

Rodgers, K., McVey, M. (2016) Error-prone repair of DNA double-strand breaks. J. Cell Physiol. 231, 15-24.

Roldán-Arjona, T., Ariza, R.R., Córdoba-Cañero, D. (2019) DNA base excision repair in plants: An unfolding story with familiar and novel characters. Front. Plant Sci. 10, 1055 (doi: 10.3389/fpls.2019.01055).

Rouillon, C., Zhou, M., Zhang, J., Politis, A., Beilsten-Edmands, V., Cannone, G., Graham, S., Robinson, C.V., Spagnolo, L., White, M.F. (2013) Structure of the CRISPR interference complex CSM reveals key similarities with cascade. Mol. Cell 52, 124–134.

Rüger, J., Ioannou, S., Castanotto, D., Stein, C.A. (2020) Oligonucleotides to the (gene) rescue: FDA approvals 2017-2019. Trends Pharmacol. Sci. 41, 27-41.

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

Saleh, A.F., Arzumanov, A.A., Gait, M.J. (2012) Overview of alternative oligonucleotide chemistries for exon skipping. Meth. Mol. Biol. 867, 365-378.

SAM (Scientific Advice Mechanism) (2017) New techniques in agricultural biotechnology. Explanatory Note 02. Publ. Off. Eur. Union, 150 pp. (doi:10.2777/574498).

Samai, P., Pyenson, N., Jiang, W., Goldberg, G.W., Hatoum-Aslan, A., Marraffini, L.A. (2015) Co-transcriptional DNA and RNA cleavage during type III CRISPR-Cas immunity. Cell 161, 1164–1174.

Sampson, T.R., Saroj, S.D., Llewellyn, A.C., Tzeng, Y.L., Weiss, D.S.A. (2013) CRISPR/Cas system mediates bacterial innate immune evasion and virulence. Nature 497, 254–257.

Sander, J.D., Dahlborg, E.J., Goodwin, M.J., ..., Joung, J.K. (2011) Selection-free zinc-finger nuclease engineering by context-dependent assembly (CoDA). Nat. Meth. 8, 67-69.

Santos, C.N., Yoshikuni, Y. (2014) Engineering complex biological systems in bacteria through recombinase-assisted genome engineering. Nat. Protoc. 9, 1320-1336.

Sattler, M.C., Carvalho, C.R., Clarindo, W.R. (2016) The polyploidy and its key role in plant breeding. Planta 243,

281-296.

Sauer, N.J., Mozoruk, J., Miller, R.B., Warburg, Z.J., Walker, K.A., Beetham, P.R., Schöpke, C.R., Gocal, G.F.W. (2016) Oligonucleotide-directed mutagenesis for precision gene editing. Plant Biotechnol. J. 14, 496-502.

Sazani, P., Gemignani, F., Kang, S.H., Maier, M.A., Manoharan, M., Persmark, M., Bortner, D., Kole, R. (2002) Systemically delivered antisense oligomers upregulate gene expression in mouse tissues. Nat. Biotechnol. 20, 1228-1233.

Schindele, A., Dorn, A., Puchta, H. (2020) CRISPR/Cas brings plant biology and breeding into the fast lane. Curr. Opin. Biotechnol. 61, 7-14.

Schneider, M.F., Wettengel, J., Hoffmann, P.C., Stafforst, T. (2014) Optimal guide RNAs for redirecting deaminase activity of hADAR1 and hADAR2 in trans. Nucl. Ac. Res. 42, e87 (doi: 10.1093/nar/gku272).

Schoch, K.M., DeVos, S.L., Miller, R.L., Chun, S.J., Norrbom, M., Wozniak, D.F., Dawson, H.N., Bennett, C.F., Rigo, F., Miller, T.M. (2016) Increased 4R-tau induces pathological changes in a human-tau mouse model. Neuron 90, 941–947.

Schoch, K.M., Miller, T.M. (2017) Antisense oligonucleotides: Translation from mouse models to human neurodegenerative diseases. Neuron 94, 1056-1070.

Schubert, D., Lechtenberg, B., Forsbach, A., Gils, M., Bahadur, S., Schmidt, R. (2004) Silencing in *Arabidopsis* T-DNA transformants: The predominant role of a gene-specific RNA sensing mechanism versus position effects. Plant Cell 16, 2561-2572.

Sedeek, K.E.M., Mahas, A., Mahfouz, M. (2019) Plant genome engineering for targeted improvement of crop traits. Front. Plant Sci. 10, 114.

Sfeir, A., Symington, L.S. (2015) Microhomology-mediated end joining: A back-up survival mechanism or dedicated pathway? Trends Biochem. Sci. 40, 701-714.

Shan, Q., Wang, Y., Chen, K., Liang, Z., Li, J., Zhang, Y., Zhang, K., Liu, J., Voytas, D.F., Zheng, X., Zhang, Y., Gao, C. (2013) Rapid and efficient gene modification in rice and *Brachypodium* using TALENs. Mol. Plant 6, 1365-1368.

Shao, J., Wang, M., Yu, G., Zhu, S., Yu, Y., Heng, B.C., Wu, J., Ye, H. (2018) Synthetic far-red light-mediated CRISPR-dCas9 device for inducing functional neuronal differentiation. Proc. Natl. Acad. Sci. U.S.A. 115, E6722-E6730.

Shen, H., Strunks, G.D., Klemann, B.J., Hooykaas, P.J., de Pater, S. (2017) CRISPR/Cas9-induced double-strand break repair in *Arabidopsis* nonhomologous end-joining mutants. G3 (Bethesda) 7, 193-202.

Shen, M.J., Wu, Y., Yang, K., Li, Y., Xu, H., Zhang, H., Li, B.-Z., Li, X., Xiao, W.-H., Zhou, X., Mitchell, L.A., Bader, J.S., Yuan, Y., Boeke, J.D. (2018) Heterozygous diploid and interspecies SCRaMbLEing. Nat. Commun. 9, 1934.

Shin, H.Y., Wang, C., Lee, H.K., Yoo, K.H., Zeng, X., Kuhns, T., Yang, C.M., Mohr, T., Liu, C., Hennighausen, L. (2017) CRISPR/Cas9 targeting events cause complex deletions and insertions at 17 sites in the mouse genome. Nat. Commun. 8, 15464.

Shmakov, S., Smargon, A., Scott, D., Cox, D., Pyzocha, N., Yan, W., Abudayyeh, O.O., Gootenberg, J.S., Makarova, K.S., Wolf, Y.I., Severinov, K., Zhang, F., Koonin, E.V. (2017) Diversity and evolution of Class 2 CRISPR-Cas systems. Nat. Rev. Microbiol. 15, 169-182.

Shukla, V.K., Doyon, Y., Miller, J.C., ..., Urnov, F.D. (2009) Precise genome modification in the crop species *Zea mays* using zinc-finger nucleases. Nature 459, 437-441.

Siddique, A.N., Nunna, S., Rajavelu, A., Zhang, Y., Jurkowska, R.Z., Reinhardt, R., Rots, M.G., Ragozin, S., Jurkowski, T.P., Jeltsch, A. (2012) Targeted methylation and gene silencing of VEGF-A in human cells by using a designed Dnmt3a–Dnmt3L single-chain fusion protein with increased DNA methylation activity. J. Mol. Biol. 425, 479–491.

Sikora, P., Chawade, A., Larsson, M., Olsson, J., Olsson, O. (2011) Mutagenesis as a tool in plant genetics, functional genomics, and breeding. Int. J. Plant Genom. 2011, 314829 (doi:10.1155/2011/314829).

Silva, G., Poirot, L., Galetto, R., Smith, J., Montoya, G., Duchateau, P., Pâques, F. (2011) Meganucleases and other tools for targeted genome engineering: Perspectives and challenges for gene therapy. Curr. Gene Ther. 11, 11-27.

Simeonov, D.R., Gowen, B.J., Boontanrat, ..., Marson, A. (2017) Discovery of stimulation-responsive immune enhancers with CRISPR activation. Nature 549, 111-115.

Simsek, D., Brunet, E., Wong, S.Y., Katyal, S., Gao, Y., McKinnon, P.J., Lou, J., Zhang, L., Li, J., Rebar, E.J., Gregory, P.D., Holmes, M.C., Jasin, M. (2011) DNA ligase III promotes alternative nonhomologous end-joining during chromosomal translocation formation. PLoS Genet. 7, e1002080.

Sinnamon, J.R., Kim, S.Y., Corson, G.M., Song, Z., Nakai, H., Adelman, J.P., Mandel, G. (2017) Site-directed RNA repair of endogenous Mecp2 RNA in neurons. Proc. Natl. Acad. Sci. U.S.A. 114, E9395-E9402.

Siva, K., Covello, G., Denti, M.A. (2014) Exon-skipping antisense oligonucleotides to correct missplicing in neurogenetic diseases. Nucl. Ac. Ther. 24, 69–86.

Slaymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X., Zhang, F. (2016) Rationally engineered Cas9 nucleases with improved specificity. Science 351, 84-88.

Smargon, A.A., Cox, D.B.T., Pyzocha, N.K., Zheng, K., Slaymaker, I.M., Gootenberg, J.S., Abudayyeh, O.A., Essletzbichler , P., Shmakov, S., Makarova, K.S., Koonin , E.V., Zhang, F. (2017) Cas13b is a type VI-B CRISPR-associated RNA-guided RNase differentially regulated by accessory proteins Csx27 and Csx28. Mol. Cell. 65, 618-630.e7.

Smith, J., Bibikova, M., Whitby, F.G., Reddy, A.R., Chandrasegaran, S., Carroll, D. (2000) Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. Nucl. Ac. Res. 28, 3361–3369.

Smith, J., Grizot, S., Arnould, S., Duclert, A., Ephinat, J.C., Chames, P., Prieto, J., Redondo, P., Blanco, F.J., Bravo, J. (2006) A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. Nucl. Ac. Res. 34, e149.

Song, R., Zhai, Q., Sun, L., Huang, E., Zhang, Y., Zhu, Y., Guo, Q., Tian, Y., Zhao, B., Lu, H. (2019) CRISPR/Cas9 genome editing technology in filamentous fungi: progress and perspective. Appl. Microbiol. Biotechnol. 103, 6919–6932.

Spilman, M., Cocozaki, A., Hale, C., Shao, Y., Ramia, N., Terns, R., Terns, M., Li, H., Stagg, S. (2013) Structure of an RNA silencing complex of the CRISPR-Cas immune system. Mol. Cell 52, 146–152.

Sprink, T., Metje, J., Hartung, F. (2015) Plant genome editing by novel tools: TALEN and other sequence specific nucleases. Curr. Opin. Biotechnol. 32, 47-53.

Staals, R.H., Agari, Y., Maki-Yonekura, S., ..., Shinkai, A. (2013) Structure and activity of the RNA-targeting Type III-B CRISPR-Cas complex of *Thermus thermophilus*. Mol. Cell 52, 135–145.

Staals, R.H., Zhu, Y., Taylor, D.W., ..., van der Oost, J. (2014) RNA targeting by the type III-A CRISPR-Cas Csm complex of *Thermus thermophilus*. Mol. Cell 56, 518–530.

Stadler, L.J. (1928) Mutations in barley induced by X-rays and radium. Science 68, 186–187.

Stafforst, T., Schneiderm M.F. (2012) An RNA-deaminase conjugate selectively repairs point mutations. Angew. Chem. Int. Ed. Engl. 51, 11166–11169.

Stephenson, M.L., Zamecnik, P.C. (1978) Inhibition of Rous sarcoma viral RNA translation by a specific oligodeoxyribonucleotide. Proc. Natl. Acad. Sci. U.S.A. 75, 285–288.

Stepper, P., Kungulowski, G, Jurkowska, R., Chandra, T., Krueger, F., Reinhardt, R., Reik, W., Jeltsch, A., Jurkowski, T.P. (2017) Efficient targeted DNA methylation with chimeric dCas9–Dnmt3a–Dnmt3L methyltransferase. Nucl. Ac. Res. 45, 1703-1713.

Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C., Doudna, J.A. (2014) DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature 507, 62-67.

Strecker J., Ladha, A., Gardner, Z., Schmid-Burgk, J.L., Makarova, K.S., Koonin, E.V., Zhang, F. (2019) RNA-guided DNA insertion with CRISPR-associated transposases. Science 365, 48-53.

Strecker, J., Ladha, A., Makarova, K.S., Koonin, E.V., Zhang, F. (2020) Response to Comment on "RNA-guided DNA insertion with CRISPR-associated transposases". Science 368, eabb2920 (doi: 10.1126/science.abb2920).

Strutt, S.C., Torrez, R.M., Kaya, E., Negrete, O.A., Doudna, J.A. (2018) RNA-dependent RNA targeting by CRISPR-Cas9. eLife 7, e32724.

Sun, N., Zhao, H.M. (2014) A single-chain TALEN architecture for genome engineering. Mol. Biosyst. 10, 446-453.

Suzuki, K., Izpisua Belmonte, J.C. (2018) *In vivo* genome editing via the HITI method as a tool for gene therapy. J. Human Genet. 63, 157–164.

Svasti, S., Suwanmanee, T., Fucharoen, S., Moulton, H.M., Nelson, M.H., Maeda, N., Smithies, O., Kole, R. (2009) RNA repair restores hemoglobin expression in IVS2-654 thalassemic mice. Proc. Natl. Acad. Sci. U.S.A. 106, 1205–1210.

Taghbalout, A., Du, M., Jillette, N., Rosikiewicz, W., Rath, A., Heinen, C.D., Li, S., Cheng, A.W. (2019) Enhanced CRISPR-based DNA demethylation by Casilio-ME-mediated RNA-guided coupling of methylcytosine oxidation and DNA repair pathways. Nat. Commun. 10, 4296.

Tamulaitis, G., Kazlauskiene, M., Manakova, E., Venclovas, C., Nwokeoji, A.O., Dickman, M.J., Horvath, P., Siksnys, V. (2014) Programmable RNA shredding by the type III-A CRISPR-Cas system of *Streptococcus thermophilus*. Mol. Cell 56, 506–517.

Tan, J., Zhang, F., Karcher, D. Bock, R. (2019) Engineering of high-precision base editors for site-specific single nucleotide replacement. Nat. Commun. 10, 439.

Tan, J., Zhao, Y., Wang, B., Hao, Y., Wang, Y., Li, Y., Luo, W., Zong, W., Li, G., Chen, S., Ma, K., Xie, X., Chen, L., Liu, Y.G., Zhu, Q. (2020) Efficient CRISPR/Cas9-based plant genomic fragment deletions by microhomology-mediated end joining. Plant Biotechnol. J. 18, 2161-2163.

Tan, M.H., Li, Q., Shanmugam, R., ..., Li, J.B. (2017) Dynamic landscape and regulation of RNA editing in mammals. Nature 550, 249–254.

Tanenbaum, M.E., Gilbert, L.A., Qi, L.S., Weismann, J.S., Vale, R.D. (2014) A protein-tagging system for signal amplification in gene expression and fluorescence imaging. Cell 159, 635–646.

Tang, X., Ren, Q, Yang, L., Bao, Y., Zhong, Z., He, Y., Liu, S., Qi, C., Liu, B., Wang, Y., Sretenovic, S., Zhang, Y., Zheng, X., Zhang, T., Qi, Y., Zhang, Y. (2019) Single transcript unit CRISPR 2.0 systems for robust Cas9 and Cas12a mediated plant genome editing. Plant Biotechnol. J. 17, 1431-1445.

Taylor, D.W., Zhu, Y., Staals, R.H., Kornfeld, J.E., Shinkai, A., van der Oost, J., Nogales, E., Doudna, J.A. (2015) Structures of the CRISPR-Cmr complex reveal mode of RNA target positioning. Science 348, 581–585.

Terns, M.P. (2018) CRISPR-Based technologies: Impact of RNA-targeting systems. Mol. Cell 72, 404-412.

Thakore, P.I., D'Ippolito, A.M., Song, L., Safi, A., Shivakumar, N.K., Kabadi, A.M., Reddy, T.E., Crawford, G.E., Gersbach, C.A. (2015) Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. Nat. Meth. 12, 1143-1149.

Thurman, R.E., Rynes, E., Humbert, R., ..., Stamatoyannopoulos, J.A. (2012) The accessible chromatin landscape of the human genome. Nature 489, 75-82.

Tian, S., Jiang, L., Cui, X., Zhang, J., Guo, S., Li, M., Zhang, H., Ren, Y., Gong, G., Zong, M., Liu, F., Chen, Q., Xu, Y. (2018) Engineering herbicide-resistant watermelon variety through CRISPR/ Cas9-mediated base-editing. Plant Cell Rep. 37, 1353–1356.

Tiricz, H., Nagy, B., Ferenc, G., Török, K., Nagy, I., Dudits, D., Ayaydin, F., (2018) Relaxed chromatin induced by histone deacetylase inhibitors improves the oligonucleotide-directed gene editing in plant cells. J. Plant Res. 131, 179–189.

Tost J. (2010) DNA methylation: An introduction to the biology and the disease-associated changes of a promising biomarker. Mol. Biotechnol. 44, 71–81.

Truong, L.N., Li, Y., Shi, L.Z., Hwang, P.Y., He, J., Wang, H., Razavian, N., Berns, M.W., Wu, X. (2013) Microhomology-mediated End Joining and Homologous Recombination share the initial end resection step to repair DNA double-strand breaks in mammalian cells. Proc. Natl. Acad. Sci. U.S.A. 110, 7720-7725.

Tsoumpra, M.K., Fukumoto, S., Matsumoto, T., Takeda, S., Wood, M.J.A., Aoki, Y. (2019) Peptide-conjugate antisense based splice-correction for Duchenne muscular dystrophy and other neuromuscular diseases. EBioMedicine 45, 630-645.

Turan, S., Bode, J. (2011) Site-specific recombinases: from tag-and-target- to tag-and-exchange-based genomic modifications. FASEB J. 25, 4088-4107.

Vad-Nielsen, J., Lin, L., Bolund, L., Nielsen, A.L., Luo, Y. (2016) Golden Gate assembly of CRISPR sgRNA expression array for simultaneously targeting multiple genes. Cell. Mol. Life Sci. 73, 4315-4325.

Vallecillo-Viejo, I.C., Liscovitch-Brauer, N., Montiel-Gonzalez, M., Eisenberg, E., Rosenthal, J.J.C. (2018) Abundant off-target edits from site-directed RNA editing can be reduced by nuclear localization of the editing enzyme. RNA Biol. 15, 104–114.

van der Wal, E., Bergsma, A.J., Pijnenburg, J.M., van der Ploeg, A.T., Pijnappel, W.W.M.P. (2017) Antisense

oligonucleotides promote exon inclusion and correct the common c.-32-13T>G *GAA* splicing variant in pompe disease. Mol. Ther. Nucl. Ac. 7, 90–100.

Varble, A., Marraffini, L.A. (2019) Three new C's for CRISPR: Collateral, Communicate, Cooperate. Trends Genet. 35, 446–456.

Vats, S., Kumawat, S., Kumar, V., Patil, G.B. Joshi, T., Sonah, H., Sharma, T.R., Desmukh, R. (2019) Genome editing in plants: Exploration of technological advancements and challenges. Cells 8, 1386.

Veillet, F., Perrot, L., Chauvin, L., Kermarrec, M.P., Guyon-Debast, A., Chauvin, J.E., Nogué, F., Mazier, M. (2019) Transgene-free genome editing in tomato and potato plants using *Agrobacterium*-mediated delivery of a CRISPR/Cas9 cytidine base editor. Int. J Mol. Sci. 20, pii-E402.

Vincent, A. (1986) TFIIIA and homologous genes. The 'finger' proteins. Nucl. Ac. Res. 14, 4385-4391.

Vítor, A.C., Huertas, P., Legube, G., de Almeida, S.F. (2020) Studying DNA double-strand break repair: An ever-growing toolbox. Front. Mol. Biosci. 7, 24.

Vogel, P., Moschref, M., Li, Q., Merkle, T., Selvasaravanan, K.D., Li, J., Stafforst, T. (2018) Efficient and precise editing of endogenous transcripts with SNAP-tagged ADARs. Nat. Meth. 15, 535–538.

Vogel, P., Schneider, M.F., Wettengel, J., Stafforst, T. (2014) Improving site-directed RNA editing *in vitro* and in cell culture by chemical modification of the guide RNA. Angew. Chem. Int. Ed. Engl. 53, 6267–6271.

Vu, V.T., Sung, Y.W., Kim, J., Doan, D.T.H., Tran, M. T., Kim, J.Y. (2019) Challenges and perspectives in homology-directed gene targeting in monocot plants. Rice 12, 95.

Walton, R.T., Christie, K.A., Whittaker, M.N., Kleinstiver, B.P. (2020) Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. Science 368, 290-296.

Wang, H., Xu, X., Nguyen, C.M., Liu, Y., Gao, Y., Lin, X., Daley, T., Kipniss, N.H., La Russa, M., Qi, L.S. (2018a) CRISPR-mediated programmable 3D genome positioning and nuclear organization. Cell 175, 1405–1417.

Wang, G., Zhao, Z., Ke, J., ..., Yoshikuni, Y. (2019b) CRAGE enables rapid activation of biosynthetic gene clusters in undomesticated bacteria. Nat. Microbiol. 4, 2498–2510.

Wang, J., Meng, X., Hu, X., Sun, T., Li, J., Wang, K., Yu, H. (2019c) xCas9 expands the scope of genome editing with reduced efficiency in rice. Plant Biotechnol. J. 17, 709-711.

Wang, M., Mao, Y., Lu, Y., Tao, X., Zhu, J. K. (2017) Multiplex gene editing in rice using the CRISPR-Cpf1 system. Mol. Plant 10, 1011-1013.

Wang, P., Zhao, F.-J., Kopittke, P.M. (2019a) Engineering crops without genome integration using nanotechnology. Trends Plant Sci. 24, 574-577.

Wang, X., Li, J., Wang, Y., Yang, B., Wei, J., Wu, J., Wang, R., Huang, X., Chen, J., Yang, L. (2018b) Efficient base editing in methylated regions with a human APOBEC3A-Cas9 fusion. Nat. Biotechnol. 36, 946-949.

Wang, X., Ma, S., Liu, Y., Lu, W., Sun, L., Zhao. P., Xia, Q. (2019d) Transcriptional repression of endogenous genes in BmE cells using CRISPRi system. Insect Biochem. Mol. Biol. 111, 103172 (https://doi.org/10.1016/j.ibmb.2019.05.007).

Wang, Y., Cheng, X., Shan, Q., Zhang, Y., Liu, J., Gao, C., Qiu, J.L. (2014) Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. Nat. Biotechnol. 32, 947-951.

Wang, Y., Liu, J., Huang, B.O., Xu, Y.M., Li, J., Huang, L.F., Lin, J., Zhang, J., Min, Q.H., Yang, W.M., and Wang, X.Z. (2015) Mechanism of alternative splicing and its regulation. Biomed. Rep. 3, 152–158.

Wang, Y., Liu, Y., Zheng, P., Sun, J., Wang, M. (2020b) Microbial base editing: A powerful emerging technology for microbial genome engineering. Trends Biotechnol. 39, 165-180.

Wang, Y., Wang, M., Zheng, T., Hou, Y., Zhang, P., Tang, T., Wei, J., Du Q. (2020a) Specificity profiling of CRISPR system reveals greatly enhanced off-target gene editing. Sci. Rep. 10, 2269.

Wang, Z., Zhou, Z.-J., Liu, D.-P., Huang, J.-D. (2006) Single-stranded oligonucleotide-mediated gene repair in mammalian cells has a mechanism distinct from homologous recombination repair. Biochem. Biophys. Res. Commun. 350, 568-573.

Watanabe, K., Breier, U., Hensel, G., Kumlehn, J., Schubert, I., Reiss, B. (2016) Stable gene replacement in barley by targeted double-strand break induction. J. Exp. Bot. 67, 1433-1445.

Webber, B.R., Lonetree, C.L., Kluesner, M.G., ..., Moriarity, B.S. (2019) Highly efficient multiplex human T cell

engineering without double-strand breaks using Cas9 base editors. Nat. Commun. 10, 5222.

Weber, N., Halpin, C., Hannah, L.C., Jez, J.M., Kough, J., Parrott, W. (2012) Editor's choice: Crop genome plasticity and its relevance to food and feed safety of genetically engineered breeding stacks. Plant Physiol. 160, 1842-1853.

Weinthal, D., Tovkach, A., Zeevi, V., Tzfira, T. (2010) Genome editing in plant cells by zinc finger nucleases. Trends Plant Sci. 15, 308-321.

Wendt, K.E., Ungerer, J., Cobb, R.E., Zhao, H., Pakrasi, H.B. (2016) CRISPR/Cas9 mediated targeted mutagenesis of the fast growing cyanobacterium *Synechococcus elongatus* UTEX 2973. Microbiol. Cell Fact. 15, 115.

Wettengel, J., Reautschnig, P., Geisler, S., Kahle, P.J., Stafforst, T. (2016) Harnessing human ADAR2 for RNA repair - Recoding a PINK1 mutation rescues mitophagy. Nucl. Ac. Res. 17, 2797-2808.

Wiedenheft, B., Sternberg, S.H., Doudna, J.A. (2012) RNA-guided genetic silencing systems in bacteria and archaea. Nature 482, 331–338.

Wierson, W.A., Simone, B.W., WareJoncas, Z., Mann, C., Welker, J.M., Kar, B., Emch, M.J., Friedberg, I., Gendron, W.A.C., Barry, M.A., Clark, K.J., Dobbs, D.L., McGrail, M.A., Ekker, S.C., Essner, J.J. (2019) Expanding the CRISPR toolbox with ErCas12a in zebrafish and human cells. CRISPR J. 2, 417-433.

Willbanks, A., Leary, M., Greenshields, M., Tyminski, C., Heerboth, S., Lapinska, K., Haskins, K., Sarkar, S. (2016) The evolution of epigenetics: From prokaryotes to humans and its biological consequences. Genet. Epigenet. 8, 25–36.

Wirth, D., Gama-Norton, L., Riemer, P., Sandhu, U., Schucht, R., Hauser, H. (2007) Road to precision: recombinase-based targeting technologies for genome engineering. Curr. Opin. Biotechnol. 18, 411-419.

Wolter, F., Schindele, P., Puchta, H. (2019) Plant breeding at the speed of light: The power of CRISPR/Cas to generate directed genetic diversity at multiple sites. BMC Plant Biol. 19, 176.

Woo, J.W., Kim, J., Kwon, S.I., Corvalán, C., Cho, S.W., Kim, H., Kim, S.G., Kim, S.T., Choe, S., Kim, J.S. (2015) DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. Nat. Biotechnol. 33, 1162-1164.

Wright, W.D., S Shah, S.S., Heyer, W.D. (2018) Homologous recombination and the repair of DNA double-strand breaks. J. Biol. Chem. 293, 10524-10535.

Wu, H., Lima, W.F., Zhang, H., Fan, A., Sun, H., Crooke, S.T. (2004) Determination of the role of the human RNase H1 in the pharmacology of DNA-like antisense drugs. J. Biol. Chem. 279, 17181–17189.

Wu, H., Liu, Q., Shi, H., ..., Lai, L. (2018) Engineering CRISPR/Cpf1 with tRNA promotes genome editing capability in mammalian systems. Cell. Mol. Life Sci. 75, 3593-3607.

Wu, J., Yin, H. (2019) Engineering guide RNA to reduce the off-target effects of CRISPR. J. Genet. Genom. 46, 523-529.

Wu, X., Zhang, Y. (2017) TET-mediated active DNA demethylation: Mechanism, function and beyond. Nat. Rev. Genet. 18, 517–534.

Wu, Y., Gao, T., Wang, X., Hu, Y., Hu, X., Hu, Z., Pang, J., Li, Z., Xue, J., Feng, M., Wu, L., Liang, D. (2014) TALE nickase mediates high efficient targeted transgene integration at the human multi-copy ribosomal DNA locus. Biochem. Biophys. Res. Commun. 446, 261-6.

Xie, J., Ge, W., Li, N., ..., Lai, L. (2019) Efficient base editing for multiple genes and loci in pigs using base editors. Nat. Commun. 10, 2852.

Yamamoto, Y., Gerbi, S.A. (2018) Making ends meet: Targeted integration of DNA fragments by genome editing. Chromosoma 127, 405-420.

Yan, W.X., Chong, S., Zhang, H., Makarova, K.S., Koonin, E.V., Cheng, D.R., Scott, D.A. (2018) Cas13d is a compact RNA-targeting type VI CRISPR effector positively modulated by a WYL-domain-containing accessory protein. Mol. Cell 70, 327–339.

Yao, X., Wang, X., Liu, J.,Hu, X., Shi, L., Shen, X., Ying, W., Sun, X., Wang, X., Huang, P., Yanga, H. (2017) CRISPR/Cas9-mediated precise targeted integration *in vivo* using a double cut donor with short homology arms. EBio Medicin 20, 19-26.

Young, A.E., Mansour, T.A., McNabb, B.R., Owen, J.R., Trott, J.F., Brown, C.T., Van Eenennaam, A.L. (2020) Genomic and phenotypic analyses of six offspring of a genome-edited hornless bull. Nat. Biotechnol. 38, 225-232.

Yu, A.M., McVey, M. (2010) Synthesis-dependent microhomology-mediated end joining accounts for multiple types of repair junctions. Nucl. Ac. Res. 38, 5706-5717.

Yu, D., Pendergraff, H., Liu, J., Kordasiewicz, H.B., Cleveland, D.W., Swayze, E.E., Lima, W.F., Crooke, S.T., Prakash, T.P., Corey, D.R. (2012) Single-stranded RNAs use RNAi to potently and allele-selectively inhibit mutant huntingtin expression. Cell 150, 895–908.

Zakharyevich, K., Ma, Y., Tang, S., Hwang, P.Y., Boiteux, S., Hunter, N. (2010) Temporally and biochemically distinct activities of Exo1 during meiosis: Double-strand break resection and resolution of double Holliday junctions. Mol. Cell 40, 1001-1015.

Zalatan, J.G., Lee, M.E., Almeida, R., Gilbert, L.A., Whitehead, E.H., La Russa, M., Tsai, J.C., Weissman, J.S., Dueber, J.E., Qi, L.S., Lim, W.A. (2015) Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. Cell 160, 339-350.

Zaman, Q.U., Li, C., Cheng, H., Hu, Q. (2019) Genome editing opens a new era of genetic improvement in polyploid crops. Crop J. 7, 141-150.

Zamecnik, P.C., Stephenson, M.L. (1978) Inhibition of Rous sarcoma virus replication and cell transformation by a specific oligodeoxynucleotide. Proc. Natl. Acad. Sci. U.S.A. 75, 280–284.

Zebec, Z., Manica, A., Zhang, J., White, M.F., Schleper, C. (2014) CRISPR-mediated targeted mRNA degradation in the archaeon *Sulfolobus solfataricus*. Nucl. Ac. Res. 42, 5280–5288.

Zetsche, B., Heidenreich, M., Mohanraju, P., Fedorova, I., Kneppers, J., DeGennaro, E.M., Winblad, N., Choudhury, S.R., Abudayyeh, O.O., Gootenberg, J.S., Wu, W.Y., Scott, D.A., Severinov, K., van der Oost, J., Zhang, F. (2017) Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. Nat. Biotechnol. 35, 31-34.

Zhang, D., Wang, Z., Wang, N., Gao, Y., Liu, Y., Ying, W., Yan, B., Zhibin, Z., Xiuyun, L., Yuzhu, D., Xiufang, O., Chunming, X., Bao, L. (2014a) Tissue culture-induced heritable genomic variation in rice, and their phenotypic implications. PLoS One 9, e96879 (doi:10.1371/journal.pone.0096879).

Zhang, F., Cong, L., Lodato, S., Kosuri, S., Church, G.M., Arlotta, P. (2011) Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. Nat. Biotechnol. 29, 149–153.

Zhang, H., Bian, Y., Gou, X., Zhu, B., Xu, C., Qi, B., Li, N., Rustgi, S., Zhou, H., Han, F., Jiang, J., von Wettstein, D., Liu, B. (2013a) Persistent whole-chromosome aneuploidy is generally associated with nascent allohexaploid wheat. Proc. Natl. Acad. Sci. U.S.A. 110, 3447-3452.

Zhang, H., Zhang, J., Wei, P., Zhang, B., Gou, F., Feng, Z., Mao, Y., Yang, L., Zhang, H., Xu, N., Zhu, J.-K. (2014b) The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation. Plant Biotechnol. J. 12, 797-807.

Zhang, L., Jia, R., Palange, N.J., Satheka, A.C., Togo, J., An, Y., Humphrey, M., Ban, L., Ji, Y., Jin, H., Feng, X., Zheng, Y. (2015a) Large genomic fragment deletions and insertions in mouse using CRISPR/Cas9. PLoS One 10, e0120396.

Zhang, N., Roberts, H.M, Van Eck, J., Martin, G.B. (2020a) Generation and molecular characterization of CRISPR/Cas9-induced mutations in 63 immunity-associated genes in tomato reveals specificity and a range of gene modifications. Front. Plant Sci. 11, 10.

Zhang, S., Guo, F., Yan, W., Dai, Z., Dong, W., Zhou, J., Zhang, W., Xin, F., Jiang, M. (2020b) Recent advances of CRISPR/Cas9-based genetic engineering and transcriptional regulation in industrial biology. Front. Bioeng. Biotechnol. 7, 459.

Zhang, T., Zheng, Q., Yi, X., An, H., Zhao, Y., Ma, S., Zhou, G. (2018) Establishing RNA virus resistance in plants by harnessing CRISPR immune system. Plant Biotechnol. J. 16, 1415-1423.

Zhang, W.W., Matlashewski, G. (2019) Single-strand annealing plays a major role in double-strand DNA break repair following CRISPR-Cas9 cleavage in *Leishmania.* mSphere 4, e00408-19.

Zhang, X., Zhu, B., Chen, L., ..., Li, D. (2020c) Dual base editor catalyzes both cytosine and adenine base conversions in human cells. Nat. Biotechnol. 38, 856-860.

Zhang, X.H., Tee, L.Y., Wang, X.G., Huang, Q.S., Yang, S.H. (2015b) Off-target effects in CRISPR/Cas9-mediated genome engineering. Mol. Ther. Nucl. Ac. 4, e264.

Zhang, Y., Zhang, F., Li, X., Baller, J.A., Qi, Y., Starker C.G., Bogdanove, A.J., Voytas, D.F. (2013b) Transcription activator-like effector nucleases enable efficient plant genome engineering. Plant Physiol. 161, 20-27.

Zhao, B., Watanabe, G., Morten, M.J., Reid, D.A., Rothenberg, E., Lieber, M.R. (2019) The essential elements for

the noncovalent association of two DNA ends during NHEJ synapsis. Nat. Commun. 10, 3588 (12 pp).

Zhao, D., Li, J., Li, S., Xin, X., Hu, M., Price, M.A., Rosser, S.J., Bi, C., Zhang, X. (2020) Glycosylase base editors enable C-to-A and C-to-G base changes. Nat. Biotechnol. 39, 35-40.

Zhao, L., Bonocora, R.P., Shub, D.A., Stoddard, B.L. (2007) The restriction fold turns to the dark side: A bacterial homing endonuclease with a PD-(D/E)-XK motif. EMBO J. 26, 2432-2442.

Zheng, Y., Lorenzo, C., Beal, P.A. (2017) DNA editing in DNA/RNA hybrids by adenosine deaminases that act on RNA. Nucl. Ac. Res. 45, 3369-3377.

Zhou, C., Sun, Y., Yan, R., Liu, Y., Zuo, E., Gu, C., Han, L., Wei, Y., Hu, X., Zeng, R., Li, Y., Zhou, H., Guo, F., Yang, H. (2019a) Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis. Nature 571, 275-278.

Zhou, J., Xin, X., He, Y., Chen, H., Li, Q., Tang, X., Zhong, Z., Deng, K., Zheng, X., Akher, S.A., Cai, G., Qi, Y., Zhang, Y. (2019b) Multiplex QTL editing of grain-related genes improves yield in elite rice varieties. Plant Cell Rep. 38, 475-485.

Zhu, Q., Yu, S., Zeng, D., Liu, H., Wang, H., Yang, Z., Xie, X., Shen, R., Tan, J., Li, H., Zhao, X., Zhang, Q., Chen, Y., Guo, J., Chen, L., Liu, Y.-G. (2017) Development of "Purple Endosperm Rice" by engineering anthocyanin biosynthesis in the endosperm with a high-efficiency transgene stacking system. Mol. Plant 10, 918-929.

Zolotarev, N., Georgiev, P., Maksimenko, O. (2019) Removal of extra sequences with I-*Sce*I in combination with CRISPR/Cas9 technique for precise gene editing in *Drosophila*. BioTechniques 66, 198-201.

Zong, Y., Song, Q., Li, C., Jin, S., Zhang, D., Wang, Y., Qiu, J.L., Gao, C. (2018) Efficient C-to-T base editing in plants using a fusion of nCas9 and human APOBEC3A. Nat. Biotechnol. 36, 950-953.

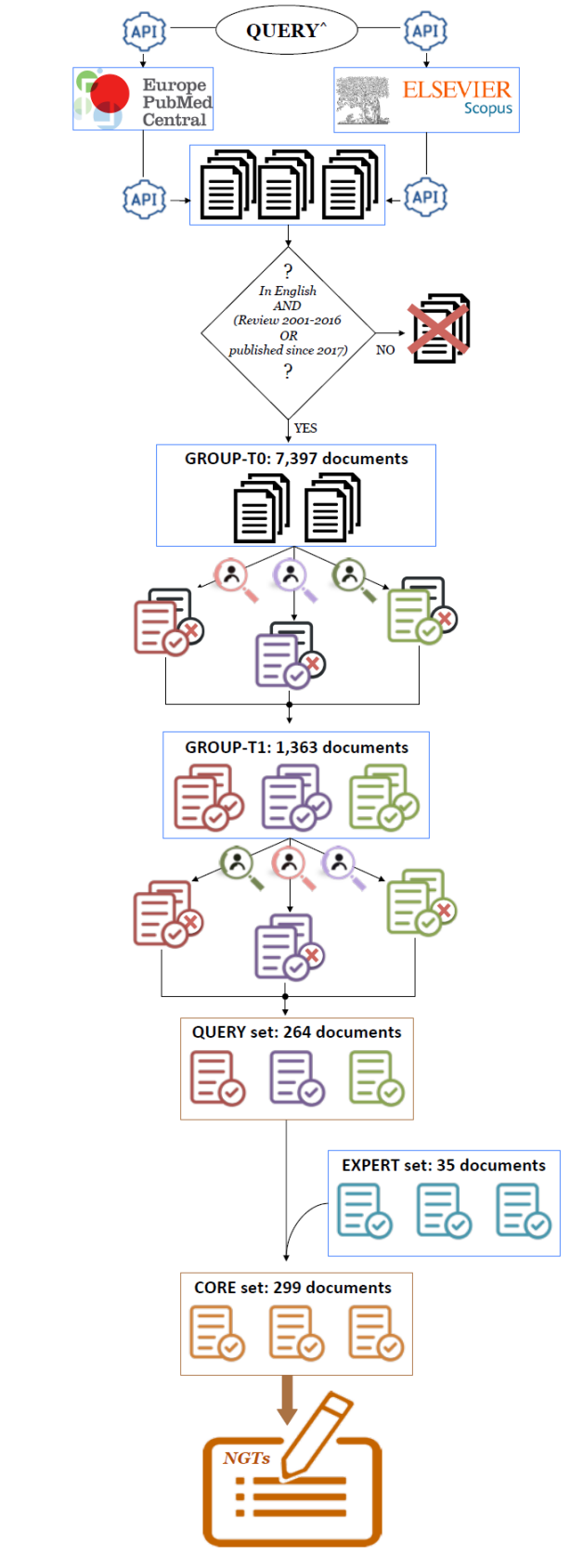
Zong, Y., Wang, Y., Li, C., Zhang, R., Chen, K., Ran, Y., Qiu, J.L., Wang, D., Gao, C. (2017) Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. Nat. Biotechnol. 35, 438-440.

Zsögön, A., Čermák, T., Naves, E.R., Notini, M.M., Edel, K.H., Weinl, S., Freschi, L., Voytas, D.F., Kudla, J., Peres, L.E.P. (2018) *De novo* domestication of wild tomato using genome editing. Nat. Biotechnol. 36, 1211-1216.

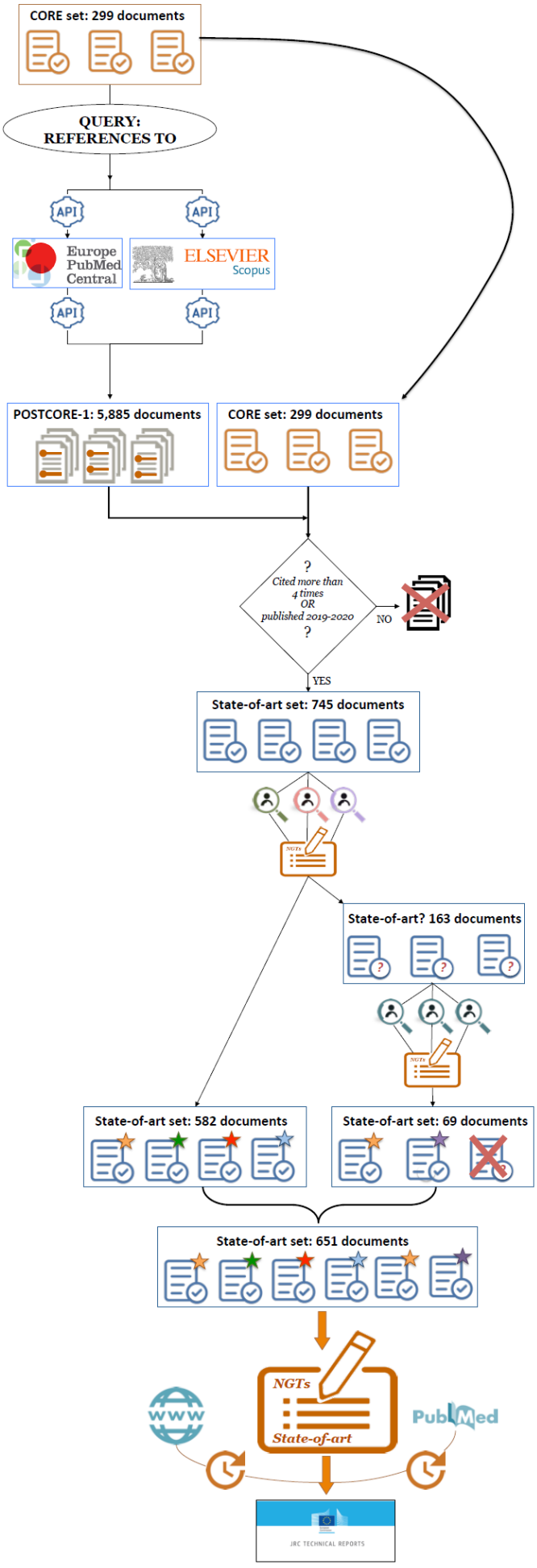
Zuo, E., Sun, Y., Wei, W., Yuan, T., Ying, W., Sun, H., Yuan, L., Steinmetz, L.M., Li, Y., Yang, H. (2019) Cytosine base editor generates substantial off-target single nucleotide variants in mouse embryos. Science 364, 289-292.

11 Annexes

Annex 1. Search strategy flowchart



^ = TITLE-ABS-KEY ((nucleic* OR genom* OR gene*) W/5 (edit* OR manipul* OR engineer* OR modif*) W/5 (technique* OR technolog*))


CORE set: 299 documents
QUERY:
REFERENCES TO
API
API
Europe PubMed Central
ELSEVIER Scopus
API
API
POSTCORE-1: 5,885 documents
CORE set: 299 documents
?
Cited more than 4 times OR published 2019-2020
?
NO
YES
State-of-art set: 745 documents
State-of-art? 163 documents
State-of-art set: 582 documents
State-of-art set: 69 documents
State-of-art set: 651 documents
NGTs
State-of-art
www
PubMed
JRC TECHNICAL REPORTS

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

- at the following standard number: +32 22999696, or

- by electronic mail via: https://europa.eu/european-union/contact_en

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications

You can download or order free and priced EU publications from EU Bookshop at: https://publications.europa.eu/en/publications. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

KJ-NA-30430-EN-N

The European Commission's science and knowledge service

Joint Research Centre

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub

ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

EU Science, Research and Innovation

EU Science Hub

Publications Office of the European Union

doi:10.2760/710056

ISBN 978-92-76-24696-1