CellPress

Review

On the Origin of CRISPR-Cas Technology: From Prokaryotes to Mammals

Francisco J.M. Mojica[1,2] and Lluis Montoliu[3,4,*]

Clustered regularly-interspaced short palindromic repeat (CRISPR) sequences cooperate with CRISPR-associated (Cas) proteins to form the basis of CRISPR-Cas adaptive immune systems in prokaryotes. For more than 20 years, these systems were of interest only to specialists, mainly molecular microbiologists, who tried to understand the properties of this unique defense mechanism. In 2012, the potential of CRISPR-Cas systems was uncovered and these were presented as genome-editing tools with an outstanding capacity to trigger targeted genetic modifications that can be applied to virtually any organism. Shortly thereafter, in early 2013, these tools were shown to efficiently drive specific modification of mammalian genomes. This review attempts to summarize, in a comprehensive manner, the key events and milestones that brought CRISPR-Cas technology from prokaryotes to mammals.

Limitations in Mammalian Genome Engineering Are Usually Overcome with Tools Imported from Prokaryotes

When researchers encounter difficulties in modifying the mammalian genome, prokaryotes very often come to the rescue, providing innovative solutions derived directly from nature. Useful reporter genes (such as *lacZ*) to identify cell types where a gene is expressed [1], robust inducible gene expression (the tetracycline system) [2] and efficient conditional mutagenesis (the cre/*loxP* system) [3] illustrate how bacteria have often been instrumental in triggering exceptional qualitative advances in our ability to modify the eukaryote genome at will, particularly that of vertebrates and notably, that of mammals.

For more than 30 years, mouse developmental and molecular biologists benefited from random and targeted animal transgenesis techniques [4], from simple methods based on the direct microinjection of DNA constructs into the pronucleus of fertilized eggs, to sophisticated, specific genetic modifications implemented with the help of **embryonic stem (ES) cells** (see Glossary) and homologous recombination approaches. These and other complementary methods were applied to the genome modification of livestock, fish, and other animal species [5].

For decades, most genome modifications in animals were based on stochastic approaches, whereby the transgene would integrate randomly into the host genome. This often led to chromosomal position effects [6] that resulted in variegation or in unexpected, variable, ectopic, and ultimately, suboptimal transgene expression patterns [7]. Although homologous gene recombination in ES cells provided a suitable solution for undesired random modification of genomes [8], the use of ES cells was limited for many years to mice [9]. It was not until the somatic cell nuclear transfer (SCNT) approach was developed that it became possible to alter

Trends

CRISPR-Cas systems, originally described as adaptive immune systems in bacteria and archaea, have become popular tools for the targeted genome modification in many organisms. The ease and robustness of CRISPR-Cas applications developed for eukaryotes originate from the numerous previous studies carried out in prokaryotes.

CRISPR-Cas tools follow the rationale described for other genome-editing nucleases, such as ZFN or TALEN. However, in contrast to those, CRISPR-Cas tools use short RNA molecules (and not protein domains) to drive the required homology to target double-strand breaks (DSBs) produced by the associated nucleases.

Most CRISPR-Cas applications derive from the CRISPR-Cas9 tools from *Streptococcus pyogenes*. Other CRISPR-Cas-like systems are being isolated and characterized from other prokaryotes, and more efficient tools could be discovered for improved genome modification.

[1]Departamento de Fisiología, Genética y Microbiología, Universidad de Alicante, Alicante, Spain
[2]Instituto Multidisciplinar para el Estudio del Medio, Universidad de Alicante, Alicante, Spain
[3]Departamento de Biología Molecular y Celular, Centro Nacional de Biotecnología (CNB-CSIC), Madrid, Spain
[4]CIBERER, Instituto de Salud Carlos III, Madrid, Spain

*Correspondence:
montoliu@cnb.csic.es (L. Montoliu).

Box 1. Rationale of Genome-Editing Nucleases

The rationale is conceptually identical for all three types of genome-editing nucleases [63]. The idea is to generate a double-strand break (DSB) in the genome produced by a DNA endonuclease enzyme (*Fok*I in the case of ZFN and TALEN, mainly Cas9 for CRISPR-Cas tools). The selected DNA sequence is targeted via the specific DNA-binding capacities of engineered zinc-finger or TALE protein domains, or by CRISPR-derived complementary small single guide RNAs (sgRNAs) or through CRISPR RNAs (crRNAs)-trans-activating crRNA (tracrRNA) duplex. The DSB is then repaired by endogenous DNA repair mechanisms, through the default non-homologous end-joining (NHEJ) route, or through the homology-directed repair (HDR) pathway if additional DNA molecules that match sequences in the vicinity of the DSB are available. Repair of the DSB leads to the introduction of small insertions or deletions (INDEL) or replacement of the target sequence by the exogenous DNA template; this depends on whether the NHEJ or the HDR pathway is triggered, and often causes gene disruption or gene editing, respectively [63]. Generally, irrespective of whether NHEJ or HDR pathways are triggered, the global term of gene (or genome) editing is applied to describe the endogenous repair mechanisms taking place immediately after the DSB.

Methods that use genome-editing nucleases evolved very quickly. The first report of an animal (a rat) whose genome was modified using ZFN was published in 2009 [77]. Two years later, TALENs were first applied for target genome modification [78]. After only two more years, in 2013, CRISPR-Cas9 tools were shown to mediate efficient targeted genome modification in mammalian cells [52,53], zebrafish [59] and mice [59,60]. This led to the thousands of publications that have accumulated since then using a CRISPR-Cas design, largely surpassing the number of studies using the previous ZFN and TALEN approaches. The closed architecture of ZFN, and the more affordable but still very laborious generation of TALENs, were rapidly replaced by the simplicity and ease of use of CRISPR-Cas9 tools [79], which explains the widespread adoption of this last approach, which is now used universally.

endogenous loci directly in nonrodent species [10]. Both the ES cell and the SCNT approaches nonetheless required laborious, lengthy protocols available only to a handful of research centers and reference laboratories. The need was therefore obvious for better, simpler, more efficient tools to target modification of animal genomes.

The introduction of genome-editing nucleases (at first, the zinc-finger nucleases, ZFN [11], followed by transcription activator-like effector nucleases, TALENs [12], and most recently, clustered regularly interspaced short palindromic repeats (CRISPR)-associated proteins, Cas [13]; see Figure 1) provided an operative solution for targeting any desired genetic modification, for specifically altering genes at will, which was welcomed and rapidly embraced by the scientific community (Box 1).

Initial CRISPR Discoveries

Why then is CRISPR-Cas the most recent and best example of tools imported from prokaryotes for application and benefit to genetic modifications in eukaryotes? What is the origin of the CRISPR-Cas technology? The following paragraphs outline the key discoveries and advances in prokaryote research since the early 1990s that permitted the launch of this disruptive technique in 2012, that is, the unexpected application of a basic research study that resulted in one of the most extraordinary breakthroughs in biotechnology (Table 1 gives a summary of milestones in CRISPR history).

The DNA repeats with dyad symmetry that would later become known as CRISPR were first reported in 1987 by Atsuo Nakata's group in Japan [14]. These regularly spaced motifs were clustered next to the *iap* gene, which encodes an aminopeptidase in *Escherichia coli* K12. A second array was found in the same genome 2 years later, and hybridization assays suggested the presence of similar sequences in very close relatives (*Shigella* and *Salmonella* species) [15]. In 1991, interspaced direct repeats (DR) were identified in strains of an evolutionarily distant group of bacteria, the *Mycobacterium tuberculosis* complex (MTBC) [16]. The DR-intervening sequences, known as spacers, were found to differ among isolates, and hence were harnessed for strain typing [17,18]. The use of DR loci as genetic markers for strain differentiation in MTBC expanded rapidly during the 1990s, and they are still used for this purpose today [19].

Glossary

Cas9: multidomain protein of Type II systems that catalyzes DSBs in target DNA guided by crRNA and in association with tracrRNA.
Clustered regularly-interspaced short palindromic repeats (CRISPR): repetitive DNA sequences present in genetic elements of bacteria and archaea, separated by unique spacers of similar length.
CRISPR-associated (Cas): proteins encoded by the *cas* genes which are usually located in close proximity to CRISPR loci, altogether forming a CRISPR-Cas system. The Cas proteins are involved in all stages of the CRISPR-Cas mechanism.
CRISPR RNA (crRNA): noncoding RNA made of a single spacer sequence and repeat fragments, generated by processing of a pre-crRNA. The crRNAs direct Cas proteins involved in interference to spacer-complementary nucleic acids.
Double-strand break (DSB): disruption of the physical continuity of a DNA molecule in which both strands in the double helix are cut, usually by a DNA endonuclease.
Embryonic stem (ES) cells: mammalian pluripotent cells established in culture and derived from the inner cell mass of blastocysts.
***Fok*I:** type II restriction enzyme derived from *Flavobacterium okeanokoites* whose DNA endonuclease domain has been engineered into ZFN and TALEN genome-editing nucleases.
Homology-directed repair (HDR): one of the two endogenous cellular mechanisms for fixing a DSB through the combined use of an additional DNA molecule, to be used as a template, with homologies to flanking sequences, and the endogenous repair machinery.
Induced pluripotent cell (iPS): mammalian pluripotent cells derived in culture from somatic cells through the activation of a limited number of reprogramming factors.
***Mycobacterium tuberculosis* complex (MTBC):** genetically related group of *Mycobacterium* species that cause tuberculosis.
Nonhomologous end-joining (NHEJ): together with HDR, the alternative endogenous cellular mechanism for repairing a DSB in the absence of a DNA template, usually resulting in small insertions and

Table 1. Milestones, Discoveries, and Achievements in the History of CRISPR-Cas Technology (1987–2013)

Year	Milestone	Refs
1987	First report of CRISPR arrays in Gram-negative bacteria	[14]
1991	First report of CRISPR arrays in Gram-positive bacteria	[16]
1993	First report of CRISPR arrays in archaea	[20]
1995	First insight on CRISPR functionality	[21]
2000	Large number of regularly spaced repeats are found in bacteria and archaea, suggesting a relevant function	[22]
2002	Regularly spaced repeats of bacteria and archaea are termed with the acronym CRISPR	[23]
2002	First identification of CRISPR-associated (*cas*) genes	[23]
2005	First identification of CRISPR spacers as homologous to sequences in bacteriophages and plasmids	[24–26]
2005	First suggestion that CRISPR-Cas systems would represent a bacterial defense mechanism	[24]
2007	First experimental demonstration that CRISPR-Cas systems are involved in acquired immunity against bacteriophages	[31]
2008	First experimental demonstration that CRISPR-Cas systems interfere with plasmid horizontal transfer, by targeting DNA	[34]
2008	First description of the role of CRISPR small RNAs (crRNA) as the guides for CRISPR interference	[33]
2008	It is anticipated that conserved sequences next to protospacers are important for CRISPR-mediated phage resistance	[37]
2010	First description of the CRISPR-Cas interference mechanism through Cas proteins cutting target DNA at precise sites	[38]
2011	Identification of trans-activating crRNAs (tracrRNAs)	[41]
2011	First successful transfer of a CRISPR-Cas system between two evolutionary distant organisms: from *Streptococcus thermophilus* to *Escherichia coli*	[80]
2012	First reports documenting functional CRISPR-Cas systems reconstructed *in vitro* and suggesting their potential application as RNA-programmable genome editing tools	[42,48]
2013	First reports demonstrating the use of CRISPR-Cas tools for efficient genome editing in mammalian cells	[52,53]
2013	First reports showing efficient genome engineering at multiple loci in mice, through the use of CRISPR-Cas tools	[59–61]

Following these first descriptions of CRISPR in Gram-negative bacteria (*E. coli*) [14,15] and Gram-positive bacteria (MTBC) [16–18], the research group of Francisco Rodríguez-Valera in Spain discovered repeat-spacer clusters in archaea [20]. They detected long stretches of these elements in chromosomal and in resident plasmid regions of strains pertaining to some *Haloferax* and *Haloarcula* species [21]. Whereas bacterial systems were not analyzed until the next decade, transcription from repeat loci [20] and the first studies to define a biological role for CRISPR [21] were reported in archaea, in 1993 and 1995, respectively, which described an incompatibility between the chromosome and recombinant multicopy plasmids with identical repeat arrays [21].

From 1996 to 1999, similar repeated elements were found in other archaea and bacteria, and in 2000 these and additional sequences in DNA databases were collected to designate a newly identified type of prokaryotic short repeats that was termed short regularly spaced repeats (SRSR) [22]. The then-rudimentary bioinformatic analyses applied to these SRSR elements indicated that they were partially palindromic and occurred in clusters, regularly interspersed by unique spacer sequences of constant length, similar to that of the repeats. This very first compilation of a large number of such SRSR in unrelated microorganisms, and the peculiarities

deletions (INDELs). NHEJ can also occur in the presence of homologous DNA templates.

Precursor-crRNA (pre-crRNA): RNA encoded by a CRISPR locus, usually encompassing the whole repeat array.

Protospacer adjacent motif (PAM): a two- to five-nucleotide conserved motif, occurring next to CRISPR targets of most CRISPR-Cas systems, which is necessary for efficient target cleavage.

Short regularly spaced repeats (SRSR): original denomination of the family of DNA repeats afterwards renamed as CRISPR.

Single-guide RNA (sgRNA): synthetic RNA molecule composed of crRNA and tracrRNA sequences merged through a linker element.

Somatic cell nuclear transfer (SCNT) approach: sophisticated reproduction procedure to generate animals from enucleated embryos reconstructed with nuclei from somatic cells.

Spacer: CRISPR-intervening sequence, typically of exogenous origin.

Trans-activating crRNA (tracrRNA): noncoding RNA of Class II CRISPR-Cas systems required for both the generation of crRNAs and target cleavage.

Transcription activator-like effector nuclease (TALEN): chimeric protein made of a designed transcription activator-like core domain, with specific DNA-binding capacity, fused with the DNA endonuclease domain from the *Fok*I restriction enzyme.

Zinc-finger nuclease (ZFN): chimeric protein made of several designed and sequence-specific zinc-finger DNA-binding domains fused with the DNA endonuclease domain from the *Fok*I restriction enzyme.

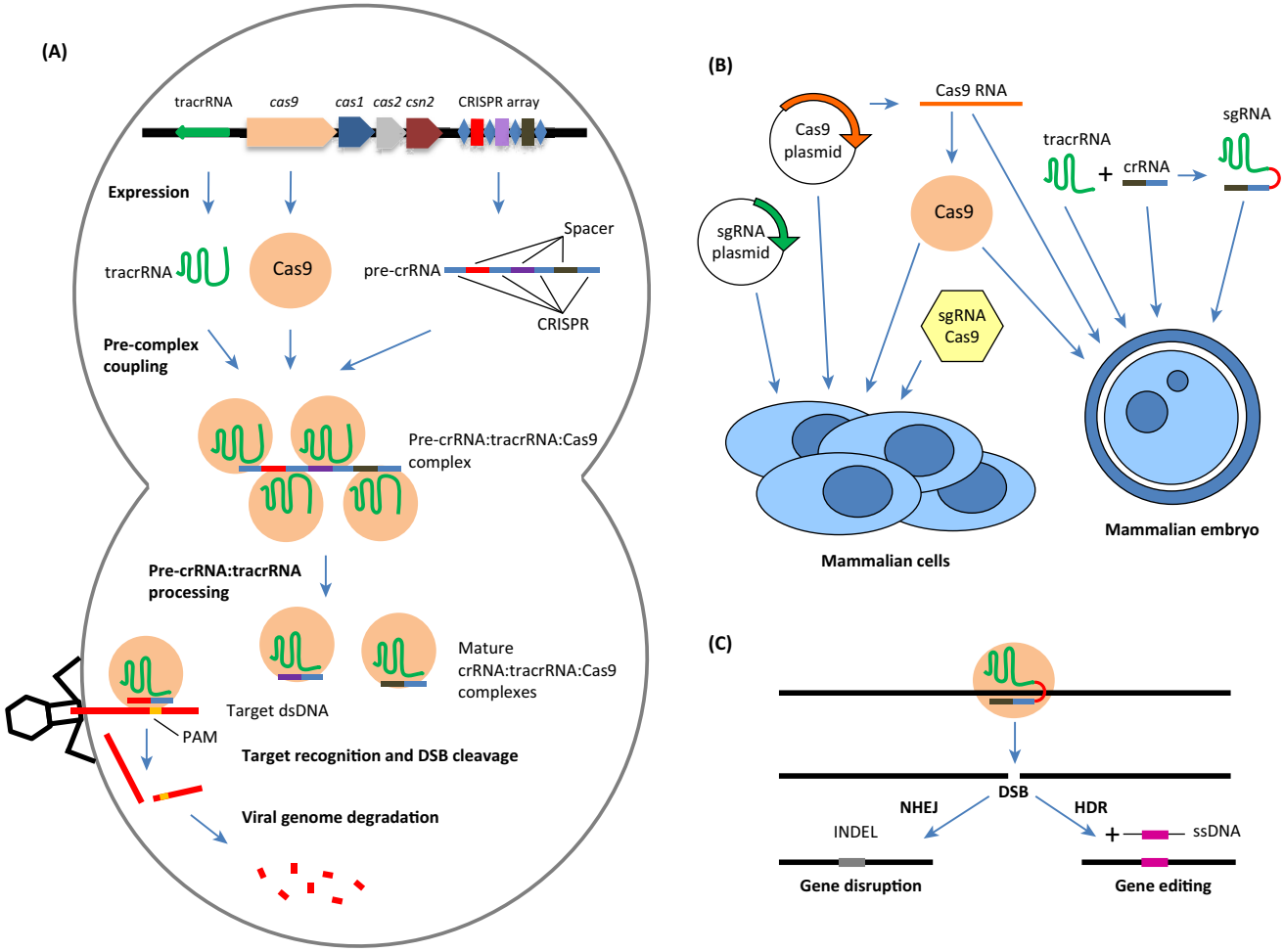

Figure 1. CRISPR-Cas Tools for Mammalian Genome Editing Derive from Prokaryotic Immune Systems. (A) The naturally occurring Type II CRISPR-Cas system in *Streptococcus pyogenes* and the mechanism of defense against viruses. The CRISPR-Cas locus is composed of four *cas* genes, a tracrRNA-encoding region and a CRISPR-spacer array. Cas9 proteins and precursor-crRNA (pre-crRNA) transcripts are bound through tracrRNA molecules in a 'pre-complex' in which RNAs are processed, through cleavage by endonuclease RNase III, into mature crRNA:tracrRNA:Cas9 complexes. When an appropriate PAM (protospacer adjacent motif; NGG in *S. pyogenes*) is detected in the genome of an invading virus, if the adjacent region matches the spacer of the associated crRNA, Cas9 catalyzes a DSB (double-strand break) that halts infection. (B) CRISPR-Cas9 tools for target genome editing in mammalian cells and embryos. Cells can be transfected with plasmids that express single-guide RNA (sgRNA) and Cas9, with Cas9 protein complexed with sgRNA (in the form of a ribonucleoprotein, RNP), or transduced with viral particles bearing sgRNA/Cas9-expressing cassettes, usually lentivirus or adeno-associated virus (AAV). Mammalian embryos can be microinjected or electroporated with Cas9 protein, Cas9 RNA, individual tracrRNA/crRNA, or sgRNA. (C) After reconstitution in the nucleus of the cell/embryo, the CRISPR-Cas9 compounds cut the DNA as a DSB, which is repaired by endogenous mechanisms. The nonhomologous end-joining (NHEJ) route is associated with the insertion and deletion (INDEL) of nucleotides, usually resulting in gene disruption. The homology-directed repair (HDR) route, in the presence of single-strand DNA (ssDNA) or double-strand DNA (dsDNA) oligonucleotides with homology to sequences surrounding the target DSB, drives repair by introducing exogenous DNA sequences, which normally leads to knock-in/reporter insertion or gene editing. Not drawn to scale.

of their genomic structure suggested an underlying, probably important but yet-unknown biological function [22].

By examining SRSR loci in many archaea and bacteria, it was possible to detect a set of four genes in their vicinity (*cas1–cas4*) that encoded proteins possibly associated with the clustered repeats [23]. This 2002 publication [23], from a team of microbiologists in the Netherlands, credited a proposal by Francisco J.M. Mojica's group to unify the diversity of names and labels used in the literature for these DNA repeated elements under the concept "clustered

regularly-interspaced short palindromic repeats", with the acronym CRISPR. This name rapidly became popular and was readily accepted in the still-incipient field.

CRISPR-Cas Systems Involved in Bacterial Immunity

Until 2005 there was no clear evidence of the biological function or the underlying mechanism associated with the CRISPR arrays. The first suggestion of a putative link between CRISPR and prokaryotic immunity arose from the discovery that some of the spacers were homologous to DNA sequences from bacteriophages or plasmids [24]. Indeed, some bacteria carrying CRISPR arrays were known to be resistant to phages, which were shown to bear these sequences [24]. This fundamental observation, independently confirmed by other laboratories [25,26], strongly supported the idea that CRISPR elements were associated with some sort of a bacterial defense system. A tentative mechanism of action was proposed [24], involving target recognition by CRISPR-RNA molecules [20,27], reminiscent of the eukaryotic interference RNA [24,28]. One of these 2005 studies [26] was also the first to observe a short stretch of conserved DNA next to the protospacers (the original sequence in viruses and other foreign genetic elements from which the spacers were derived [29]), which would later be termed the protospacer adjacent motif (PAM) [30].

CRISPR element involvement in bacterial immunity was confirmed experimentally in 2007 by Rodolphe Barrangou, Philippe Horvath and collaborators [31]. This landmark publication demonstrated that resistance to bacteriophage infection can be developed by a sensitive bacterial strain through acquisition of spacers that match the viral genome [31]. CRISPR arrays in prokaryotes had been previously found to produce numerous small RNAs [20,27,32]. In 2008, Brouns *et al*. [33] showed that these CRISPR RNAs (crRNAs) have a crucial role in driving antiviral defense. Moreover, their results suggested that DNA was the target of CRISPR action. Indeed, Luciano Marraffini and Erik Sontheimer subsequently reported that CRISPR could efficiently interfere with the horizontal transfer of plasmid sequences through DNA targeting [34]. The observation that there is a dynamic interplay between the spacer content of CRISPR arrays and potential targets in natural microbial communities further supported a connection between CRISPR genotype and host immunity [35]. These findings corroborated CRISPR-Cas as a general genetic barrier to horizontally transferred DNA and an efficient adaptive immune system in prokaryotes [36].

Over the following years, these seminal discoveries allowed further detailed characterization of the CRISPR mechanism of action. PAM sequences appeared to be important for interference [37], and these motifs emerged as a common feature in many systems [29,30], further supporting their functional relevance. The interference/defense mechanism was then defined as Cas protein cleavage of target DNA next to a PAM [38]. However, some systems were documented to cleave RNA instead, not requiring a specific sequence motif next to the target [39,40]. Another crucial piece of information of CRISPR-Cas was uncovered in 2011 [41], the existence in particular systems of an additional small RNA molecule, the trans-activating crRNA (tracrRNA), which was needed to generate mature crRNA molecules. TracrRNAs are exclusive of some Class 2 systems (see below), where, in addition to their implication in crRNA maturation, they also bridge crRNA and the Cas protein responsible for target cleavage [42]. By that time, many CRISPR-Cas systems had been identified [43] and partially characterized in archaea and bacteria [44] (Box 2). This large amount of information on CRISPR-Cas systems led to the first attempt to classify them, from an evolutionary perspective, into distinct functional and structural types (Type I, II, and III) and subtypes [45]. A top level classification, the Class category (that is, Class 1, comprising types I, III, and IV, and Class 2, including types II and V), has recently been adopted [46]. In contrast to Class 1, Class 2 systems require only one Cas protein (Cas9 in the case of Type II systems, instead of a multiprotein complex as in Class 1) for target recognition and cleavage, producing single DNA cuts [47]. These properties of Class 2 systems explain why

Box 2. First Successful Attempt Transferring a CRISPR-Cas System between Evolutionary Distant Organisms

Until 2011, the CRISPR-Cas systems had always been studied *in vitro* or using their endogenous hosts. At the end of that year, a collaborative effort in the already very active CRISPR field, led by Virginijus Siksnys in Lithuania, provided the first experimental demonstration that one such CRISPR-Cas system could be transferred between the evolutionarily very distant bacterial species *Streptococcus thermophilus* and *Escherichia coli*, with transmission of the immunity developed by the former bacterium into the latter [80]. This was an outstanding achievement, as the associated evolutionary distance between these two species is greater than that between humans and yeast [81,82]. This experiment illustrated the broad permissiveness of CRISPR-Cas systems and proved that CRISPR arrays could be reprogrammed to target novel sequences and operate across species as self-contained units [80].

Type II were chosen among characterized CRISPR-Cas systems for development of future applications based on target cleavage.

CRISPR-Cas Introduced as New Genome-Editing Tools

In the summer of 2012 two independent teams reported the biochemical properties of partially reconstituted CRISPR-Cas systems *in vitro*, and went beyond the state-of-the-art knowledge by suggesting that its elements could be used as genome-editing tools [42,48]. In a combined effort, the laboratories of Jennifer Doudna in the USA and Emmanuelle Charpentier in Sweden reconstituted *in vitro* and demonstrated the function of three of the six elements of the *Streptococcus pyogenes* CRISPR-Cas9 system (tracrRNA, crRNA and Cas9 protein; see Figure 1A), which target a double-strand break (DSB) and cut a specific DNA sequence homologous to the crRNA spacer region [42]. The same study also showed how tracrRNA and crRNA could be fused into a sgRNA, a chimeric synthetic RNA molecule that retains the full properties of the two original small RNAs. This remarkable achievement further simplified an already very simple RNA-directed DNA endonuclease mechanism for use as a tool in programmable genome editing [42].

In parallel to the Doudna and Charpentier study [42,49], the Siksnys group collaborated with Barrangou and Horvath to assess the function of the *Streptococcus thermophilus* Type II system *in vitro* [48]. Their results were similar, and demonstrated the crucial role of crRNA and Cas9 complexes in directing DSB in crRNA-targeted DNA sequences. They also understood the relevance of these findings, and proposed that universal programmable RNA-guided DNA endonucleases could be engineered as unique molecular tools for RNA-directed DNA surgery [48]. These two inspiring publications triggered a few laboratories to assess the conjectured genome-editing capacities of these newly characterized bacteria-derived, RNA-programmable DNA endonucleases [50].

It took less than 6 months to experimentally confirm the predictions of these founder publications of the nascent field of CRISPR-Cas9 technology. By January 2013, three independent US teams, one led by Luciano Marraffini [51], another by Feng Zhang [52], in collaboration with Luciano Marraffini, and a third one by George Church [53], communicated the successful editing of bacterial [51] and mammalian genomes [52,53] using Cas9. CRISPR-Cas9 tools derived from *S. pyogenes* were improved, and the *cas9* gene adapted to mammals by human codon-optimization, for efficient genome modification of various mammalian cell types from mice and humans [52,53], including pluripotent cells. Later the same month, an independent study also reported the formation of DSB at a specific locus, in human cells, using CRISPR-Cas9 methods [54]. The current CRISPR excitement had begun and the rest of the scientific community, including many who had probably not noticed the two 2012 publications on *in vitro* studies, learned of these far-reaching tools for genome editing in eukaryotes [55–57] and prokaryotes [57,58].

It took a few more months, still in 2013, before the first publications appeared that reported similar *in vivo* findings in vertebrates. A team in China presented preliminary results on genome

Box 3. The Successful Application of CRISPR-Cas Tools for Mammalian Genome Editing

In mammals, CRISPR-Cas9 tools have now been applied successfully in numerous projects:

- To functionally assess mutations in coding [83,84] and noncoding [85,86] mouse genomic DNA sequences.
- To generate genome-edited nonhuman primates [87].
- To explore genome editing in nonviable human embryos [88].
- To alter epigenetic markers leading to the activation of specific genes [89].
- To generate genome-edited livestock for biomedical [90] and agricultural applications [91].
- To inactivate multiple retroviral genomic insertions simultaneously in porcine cells, for xenotransplation purposes [92].
- To assess gene and cell therapy approaches for human diseases [93].
- To develop lentiviral knockout libraries that facilitate forward genetics in human cells [94–96].
- To reproduce chromosomal rearrangements, large inversions, and translocations observed in patients that were extremely difficult to model in human cells [97] and in mice [98].
- To apply rapid somatic genome-editing approaches for cancer modeling [99].
- To correct the genome in patient-derived induced pluripotent cells (iPS) [100,101].
- To develop *in vivo* genome-editing protocols [67] and promising somatic gene therapy approaches for devastating degenerative [102–104] or rare [105,106] diseases in animal models.

Outstanding Questions

Are there further CRISPR-Cas and CRISPR-like systems in prokaryotic genetic elements that could be harnessed for genome editing? If so, will they allow for a notable improvement of this technology?

Do other yet unknown mechanisms exist in bacteria and archaea that could be applied to modify the eukaryotic genome?

Can the engineered CRISPR-Cas tools be further improved for more precise mammalian genome editing?

Can we envisage more applications of CRISPR-Cas components in eukaryotic cells?

Will the CRISPR-Cas technology soon be reasonably safe to be applied in humans to cure genetic disorders?

Will society and governments reach a consensus, with the help of researchers, on the ethics and regulatory considerations of the use of CRISPR-Cas technology for gene therapy and environmental applications?

editing in mouse and zebrafish embryos, although they did not provide information regarding germ-line transmission of the mutations induced by CRISPR-Cas-mediated genome editing [59]. CRISPR-Cas9 was officially presented as the newest tool for improved targeted mammalian transgenesis in May 2013, when Rudolf Jaenisch's laboratory, in collaboration with Feng Zhang, published their impressive study introducing multiple mutations in specific genes in a single step as well as simultaneous editing of two genes through direct delivery of CRISPR-Cas9 reagents into mouse embryos or ES cells [60]. Later that year, the same laboratory further documented the versatility of CRISPR-Cas9 approaches for mouse genome editing by showing its application in producing a knock-in reporter and conditional mutant alleles [61]. These four pioneer articles [52,53,60,61], together with the deposition and efficient academic dissemination of all described CRISPR-Cas9 reagents through the non-profit plasmid repository Addgene [62], launched a myriad of *in vitro* studies in many cell types and *in vivo* experiments in many mammalian species [55,56,63] (Figure 1B, Box 3).

Concluding Remarks

Researchers have been astonished to confirm the ease, efficiency, and apparently unlimited number of applications that arise from the use of CRISPR-Cas components. It is important, however, that we not forget the origin of these tools. CRISPR-Cas systems have been evolving in bacteria and archaea subject to strong selection by infectious genetic elements for billions of years. It should hence not be a surprise to discover that, after this long period of optimization of tools meant to cleave intruder DNA, these elements would also perform most efficiently for genome editing outside their natural context. That which today benefits mammals is derived from what once evolved in archaea and bacteria. Likewise, the reagents we now apply for genome editing in mammals could not be understood without the systematic, fundamental, and often underestimated contribution of the many microbiologists who discovered, dissected and described the functional compounds of the native CRISPR-Cas systems [47,64].

The characterization of other known and to be discovered CRISPR-Cas systems from different prokaryotes [46,65,66] ensures a regular flow of new reagents with slightly different and useful properties [67,68]. Notably, this growing list includes Cpf1 [69], a biochemically validated nuclease of Type V CRISPR-Cas systems [46]. Compared to Cas9, Cpf1 requires a different, T-rich PAM sequence at the 5′ location, and interacts with a single shorter RNA molecule, producing PAM-distal protruding ends upon cutting the DNA [69]. Together with the generation of improved mutant variants of currently known Cas proteins [70–72], through the structure-guided rational design of Cas9 PAM variants [73,74], and with the description of similar gene-editing properties from unrelated immune systems of prokaryotes (i.e., prokaryotic Argonaute,

pAgo [75,76]), all these approaches will further widen our capacity to edit complex genomes at will (see Outstanding Questions).

Acknowledgments

The authors wish to thank Luciano Marraffini and Davide Seruggia for critical reading of this manuscript and Catherine Mark for editorial assistance. CRISPR-Cas research in FJMM and LM laboratories is funded through the Spanish Ministry of Economy and Competitiveness (MINECO) [BIO2012-39980 and BIO2015-70978-R] to LM, and [BIO2014-53029-P] to FJMM; the Centre for Networked Biomedical Research on Rare Diseases (CIBERER-ISCIII) [ACCI 2015] to LM, and the Biomedical and Biological Sciences (BMBS) European Cooperation in Science and Technology (COST) action [BM1308 SALAAM] to LM. FJMM and LM conceived, discussed and wrote the manuscript.

References

1. Gossler, A. *et al.* (1989) Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes. *Science* 244, 463–465
2. Gossen, M. and Bujard, H. (1992) Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc. Natl. Acad. Sci. U.S.A.* 89, 5547–5551
3. Gu, H. *et al.* (1994) Deletion of a DNA polymerase beta gene segment in T cells using cell type-specific gene targeting. *Science* 265, 103–106
4. Pease, S. and Saunders, T.L., eds (2011) *Advanced Protocols for Animal Transgenesis. An ISTT Manual*, Springer- Verlag
5. Niemann, H. and Kues, W.A. (2003) Application of transgenesis in livestock for agriculture and biomedicine. *Anim. Reprod. Sci.* 79, 291–317
6. Bishop, J.O. (1996) Chromosomal insertion of foreign DNA. *Reprod. Nutr. Dev.* 36, 607–618
7. Montoliu, L. (2002) Gene transfer strategies in animal transgenesis. *Cloning Stem Cells* 4, 39–46
8. Capecchi, M.R. (2005) Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. *Nat. Rev. Genet.* 6, 507–512
9. Martello, G. and Smith, A. (2014) The nature of embryonic stem cells. *Annu. Rev. Cell. Dev. Biol.* 30, 647–675
10. McCreath, K.J. *et al.* (2000) Production of gene-targeted sheep by nuclear transfer from cultured somatic cells. *Nature* 405, 1066–1069
11. Rémy, S. *et al.* (2010) Zinc-finger nucleases: a powerful tool for genetic engineering of animals. *Transgenic Res.* 19, 363–371
12. Sommer, D. *et al.* (2015) TALEN-mediated genome engineering to generate targeted mice. *Chromosome Res.* 23, 43–55
13. Seruggia, D. and Montoliu, L. (2014) The new CRISPR-Cas system: RNA-guided genome engineering to efficiently produce any desired genetic alteration in animals. *Transgenic Res.* 23, 707–716
14. Ishino, Y. *et al.* (1987) Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J. Bacteriol.* 169, 5429–5433
15. Nakata, A. *et al.* (1989) Unusual nucleotide arrangement with repeated sequences in the *Escherichia coli* K-12 chromosome. *J. Bacteriol.* 171, 3553–3556
16. Hermans, P.W. *et al.* (1991) Insertion element IS987 from *Mycobacterium bovis* BCG is located in a hot-spot integration region for insertion elements in *Mycobacterium tuberculosis* complex strains. *Infect. Immun.* 59, 2695–2705
17. Jeffreys, A.J. *et al.* (1991) Minisatellite repeat coding as a digital approach to DNA typing. *Nature* 354, 204–209
18. Groenen, P.M. *et al.* (1993) Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*; application for strain differentiation by a novel typing method. *Mol. Microbiol.* 10, 1057–1065
19. Botelho, A. *et al.* (2015) Clustered regularly interspaced short palindromic repeats (CRISPRs) analysis of members of the Mycobacterium tuberculosis complex. *Methods Mol. Biol.* 1247, 373–389
20. Mojica, F.J. *et al.* (1993) Transcription at different salinities of *Haloferax mediterranei* sequences adjacent to partially modified PstI sites. *Mol. Microbiol.* 9, 613–621
21. Mojica, F.J. *et al.* (1995) Long stretches of short tandem repeats are present in the largest replicons of the Archaea *Haloferax mediterranei* and *Haloferax volcanii* and could be involved in replicon partitioning. *Mol. Microbiol.* 17, 85–93
22. Mojica, F.J. *et al.* (2000) Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol. Microbiol.* 36, 244–246
23. Jansen, R. *et al.* (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* 43, 1565–1575
24. Mojica, F.J. *et al.* (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* 60, 174–182
25. Pourcel, C. *et al.* (2005) CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* 151, 653–663
26. Bolotin, A. *et al.* (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* 151, 2551–2561
27. Tang, T.H. *et al.* (2002) Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *Proc. Natl. Acad. Sci. U.S.A.* 99, 7536–7541
28. Makarova, K.S. *et al.* (2006) A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology Direct* 1, 7
29. Horvath, P. *et al.* (2008) Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J. Bacteriol.* 190, 1401–1412
30. Mojica, F.J. *et al.* (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* 155, 733–740
31. Barrangou, R. *et al.* (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315, 1709–1712
32. Lillestøl, R.K. *et al.* (2006) A putative viral defence mechanism in archaeal cells. *Archaea* 2, 59–72
33. Brouns, S.J. *et al.* (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321, 960–964
34. Marraffini, L.A. and Sontheimer, E.J. (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* 322, 1843–1845
35. Andersson, A.F. and Banfield, J.F. (2008) Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* 320, 1047–1050
36. Marraffini, L.A. (2015) CRISPR-Cas immunity in prokaryotes. *Nature* 526, 55–61
37. Deveau, H. *et al.* (2008) Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.* 190, 1390–1400
38. Garneau, J.E. *et al.* (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* 468, 67–71

39. Hale, C.R. *et al.* (2009) RNA-Guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* 139, 945–956
40. Hale, C.R. *et al.* (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
41. Deltcheva, E. *et al.* (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* 471, 602–607
42. Jinek, M. *et al.* (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821
43. Haft, D.H. *et al.* (2005) A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput. Biol.* 1, e60
44. Horvath, P. and Barrangou, R. (2010) CRISPR/Cas, the immune system of bacteria and archaea. *Science* 327, 167–170
45. Makarova, K.S. *et al.* (2011) Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.* 9, 467–477
46. Makarova, K.S. *et al.* (2015) An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* 13, 722–736
47. van der Oost, J. *et al.* (2014) Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nat. Rev. Microbiol.* 12, 479–492
48. Gasiunas, G. *et al.* (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 109, E2579–E2586
49. Lander, E.S. (2016) The Heroes of CRISPR. *Cell* 164, 18–28
50. Barrangou, R. (2012) RNA-mediated programmable DNA cleavage. *Nat. Biotechnol.* 30, 836–838
51. Jiang, W. *et al.* (2013) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol.* 31, 233–239
52. Cong, L. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823
53. Mali, P. *et al.* (2013) RNA-guided human genome engineering via Cas9. *Science* 339, 823–826
54. Jinek, M. *et al.* (2013) RNA-programmed genome editing in human cells. *Elife* 2, e00471
55. Doudna, J.A. and Charpentier, E. (2014) Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* 346, 1258096
56. Hsu, P.D. *et al.* (2014) Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 157, 1262–1278
57. Gasiunas, G. and Siksnys, V. (2013) RNA-dependent DNA endonuclease Cas9 of the CRISPR system: Holy Grail of genome editing? *Trends Microbiol.* 21, 562–567
58. Selle, K. and Barrangou, R. (2015) Harnessing CRISPR-Cas systems for bacterial genome editing. *Trends Microbiol.* 23, 225–232
59. Shen, B. *et al.* (2013) Generation of gene-modified mice via Cas9/RNA-mediated gene targeting. *Cell Res.* 23, 720–723
60. Wang, H. *et al.* (2013) One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* 153, 910–918
61. Yang, H. *et al.* (2013) One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell* 154, 1370–1379
62. Joung, J.K. *et al.* (2015) Accelerating research through reagent repositories: the genome editing example. *Genome Biol.* 16, 255
63. Sander, J.D. and Joung, J.K. (2014) CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* 32, 347–355
64. Barrangou, R. and van der Oost, J., eds (2013) *CRISPR-Cas Systems. RNA-Mediated Adaptive Immunity in Bacteria and Archaea*, Springer- Verlag
65. Shmakov, S. *et al.* (2015) Discovery and functional characterization of diverse class 2 CRISPR-Cas systems. *Mol. Cell.* 60, 385–397
66. Mougiakos, I. *et al.* (2016) Next generation prokaryotic engineering: the CRISPR-Cas toolkit. *Trends Biotechnol.* Published online March 1, 2016. http://dx.doi.org/10.1016/j.tibtech.2016.02.004
67. Ran, F.A. *et al.* (2015) *In vivo* genome editing using *Staphylococcus aureus* Cas9. *Nature* 520, 186–191
68. Karvelis, T. *et al.* (2015) Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements. *Genome Biol.* 16, 253
69. Zetsche, B. *et al.* (2015) Cpf1 Is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 163, 759–771
70. Kleinstiver, B.P. *et al.* (2015) Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* 523, 481–485
71. Slaymaker, I.M. *et al.* (2016) Rationally engineered Cas9 nucleases with improved specificity. *Science* 351, 84–88
72. Kleinstiver, B.P. *et al.* (2016) High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* 529, 490–495
73. Anders, C. *et al.* (2016) Structural plasticity of PAM recognition by engineered variants of the RNA-guided endonuclease Cas9. *Mol. Cell* 61, 895–902
74. Hirano, H. *et al.* (2016) Structure and engineering of *Francisella novicida* Cas9. *Cell* 164, 950–961
75. Swarts, D.C. *et al.* (2016) DNA-guided DNA interference by a prokaryotic Argonaute. *Nature* 507, 258–261
76. Gao, F. *et al.* (2016) DNA-guided genome editing using the *Natronobacterium gregoryi* Argonaute. *Nat. Biotechnol* Published online May 2, 2016. http://dx.doi.org/10.1038/nbt.3547
77. Geurts, A.M. *et al.* (2009) Knockout rats via embryo microinjection of zinc-finger nucleases. *Science* 325, 433
78. Miller, J.C. *et al.* (2011) A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.* 29, 143–148
79. Harms, D.W. *et al.* (2014) Mouse genome editing using the CRISPR/Cas system. *Curr. Protoc. Hum. Genet.* 83, 15 7.1-15.7.27
80. Sapranauskas, R. *et al.* (2011) The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res.* 39, 9275–9282
81. Segata, N. and Huttenhower, C. (2011) Toward an efficient method of identifying core genes for evolutionary and functional microbial phylogenies. *PLoS ONE* 6, e24704
82. Kachroo, A.H. *et al.* (2015) Evolution. Systematic humanization of yeast genes reveals conserved functions and genetic modularity. *Science* 348, 921–925
83. Shen, B. *et al.* (2014) Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat. Methods* 11, 399–402
84. Yang, H. *et al.* (2014) Generating genetically modified mice using CRISPR/Cas-mediated genome engineering. *Nat. Protoc.* 9, 1956–1968
85. Seruggia, D. *et al.* (2015) Functional validation of mouse tyrosinase non-coding regulatory DNA elements by CRISPR-Cas9-mediated mutagenesis. *Nucleic Acids Res.* 43, 4855–4867
86. Guo, Y. *et al.* (2015) CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell* 162, 900–910
87. Niu, Y. *et al.* (2014) Generation of gene-modified cynomolgus monkey via Cas9/RNA-mediated gene targeting in one-cell embryos. *Cell* 156, 836–843
88. Liang, P. *et al.* (2015) CRISPR/Cas9-mediated gene editing in human tripronuclear zygotes. *Protein Cell* 6, 363–372
89. Hilton, I.B. *et al.* (2015) Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat. Biotechnol.* 33, 510–517
90. Peng, J. *et al.* (2015) Production of human albumin in pigs through CRISPR/Cas9-mediated knockin of human cDNA into swine albumin locus in the zygotes. *Sci. Rep.* 5, 16705
91. Crispo, M. *et al.* (2015) Efficient generation of myostatin knockout sheep using CRISPR/Cas9 technology and microinjection into zygotes. *PLoS ONE* 10, e0136690
92. Yang, L. *et al.* (2015) Genome-wide inactivation of porcine endogenous retroviruses (PERVs). *Science* 350, 1101–1104
93. Chang, C.W. *et al.* (2015) Modeling human severe combined immunodeficiency and correction by CRISPR/Cas9-enhanced gene targeting. *Cell Rep.* 12, 1668–1677

94. Wang, T. *et al.* (2014) Genetic screens in human cells using the CRISPR-Cas9 system. *Science* 343, 80–84
95. Shalem, O. *et al.* (2014) Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* 343, 84–87
96. Konermann, S. *et al.* (2015) Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* 517, 583–588
97. Torres, R. *et al.* (2014) Engineering human tumour-associated chromosomal translocations with the RNA-guided CRISPR-Cas9 system. *Nat. Commun.* 5, 3964
98. Maddalo, D. *et al.* (2014) *In vivo* engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. *Nature* 516, 423–427
99. Sánchez-Rivera, F.J. *et al.* (2014) Rapid modelling of cooperating genetic events in cancer through somatic genome editing. *Nature* 516, 428–431
100. Huang, X. *et al.* (2015) Production of gene-corrected adult beta globin protein in human erythrocytes differentiated from patient iPSCs after genome editing of the sickle point mutation. *Stem Cells* 33, 1470–1479
101. Park, C.Y. *et al.* (2015) Functional correction of large factor viii gene chromosomal inversions in hemophilia a patient-derived iPSCs using CRISPR-Cas9. *Cell Stem Cell* 17, 213–220
102. Long, C. *et al.* (2016) Postnatal genome editing partially restores dystrophin expression in a mouse model of muscular dystrophy. *Science* 351, 400–403
103. Nelson, C.E. *et al.* (2016) *In vivo* genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy. *Science* 351, 403–407
104. Tabebordbar, M. *et al.* (2016) *In vivo* gene editing in dystrophic mouse muscle and muscle stem cells. *Science* 351, 407–411
105. Yin, H. *et al.* (2016) Therapeutic genome editing by combined viral and non-viral delivery of CRISPR system components *in vivo*. *Nat. Biotechnol.* 34, 328–333
106. Yang, Y. *et al.* (2016) A dual AAV system enables the Cas9-mediated correction of a metabolic liver disease in newborn mice. *Nat. Biotechnol.* 34, 334–338