Genome Editing Revolution in Life Sciences*

Alok Kumar Singh, Sivaprakash Ramalingam, Desirazu N Rao and Srinivasan Chandrasegaran

Programmable nucleases—ZFNs, TALENs and CRISPR-Cas9 —have equipped scientists with an unprecedented ability to modify cells and organisms almost at will, with great implications across life sciences: biology, agriculture, ecology and medicine. Nucleases-based genome editing (aka gene editing) depends on cellular responses to a targeted double-strand break (DSB). The first truly targetable reagents were zinc finger nucleases (ZFNs) showing that arbitrary DNA sequences within a mammalian genome, could be addressed by protein engineering, ushering in the era of genome editing. ZFNs that are fusions of zinc finger proteins (ZFPs) and FokI cleavage domain, resulted from the basic research on Type IIs FokI restriction enzyme, which showed a bipartite structure with a separable DNA-binding domain and a non-specific cleavage domain. Studies on 3-finger ZFNs established that the preferred substrates were paired binding sites, which doubled the size of the target recognition sequence from 9 to 18 bp that is large enough to specify a unique genomic locus in plant and mammalian cells, including human cells. Subsequently, a ZFN-induced DSB was shown to stimulate homologous recombination in frog eggs. Transcription activator-like effector nucleases (TALENs) that are based on bacterial TALEs fused to FokI cleavage domain expanded the capability. ZFNs and TALENs have been successfully used to modify a multitude of recalcitrant organisms and cell types that were unapproachable previously attesting to the success of protein engineering, long before the arrival of CRISPR. The recent technique to deliver a targeted DSB to cellular genomes are RNA-guided nucleases as exemplified by the Type II prokaryotic

Alok Kumar Singh
Interested in bacterial immunotherapy and cancer immunology.

Sivaprakash Ramalingam
Interested in therapeutic genome engineering in hematological diseases.

Desirazu N Rao
Interested in DNA-protein interactions using restriction-modification enzymes and DNA mismatch repair proteins as model systems.

Srinivasan Chandrasegaran
Interested in genome editing and synthetic biology.

CRISPR-Cas9 system. Unlike ZFNs and TALENs that use protein motifs for DNA sequence recognition, CRISPR-Cas9 depends on RNA-DNA recognition. The advantages of the CRISPR-Cas9 system, which include ease of RNA design for new targets and dependence on a single constant Cas9 protein, have led to its wide adoption by research labs around the world. The 2020 Nobel Prize for Chemistry was awarded to Jennifer Doudna and Emmanualle Charpentier for harnessing CRISPR-Cas9 system to provide a simplified technique for genome editing. The programmable nucleases have also been shown to cut at off-target sites with mutagenic consequences, which is a serious concern for human therapeutic applications. Therefore, applications of genome editing technologies to human therapeutics will ultimately depend on risk versus benefit analysis and informed consent

Introduction

The nucleotide sequence of an organism's genome encodes its form and function. For example, each human cell contains 3 billion base pairs (3×10^9 bp) within its genome. Scientists estimate that the human genome codes for 20,000 to 25,000 different genes, which gives an individual form and function. The genes are distributed among the 23 pairs of chromosomes and code for all of the human body's proteins. While the functions of several human genes have been deciphered, many others are yet to be determined. The gene function, in many cases, can be inferred by how well the gene and its protein sequence are conserved across species, and if the gene function of one of the homologs is already known. The function of a gene can be determined experimentally by knocking out (deleting) or mutating that specific gene individually in a cell and observing the resulting phenotype. Mutations in the coding region of critical genes can lead to the formation of abnormal proteins, result in disease phenotypes, premature death, or failure of an embryo to develop. Moreover, mutations that affect the regulatory genes can result in aberrant gene expression within cells and give rise to cancer phenotypes. The 'holy grail'

> The gene function, in many cases, can be inferred by how well the gene and its protein sequence are conserved across species, and if the gene function of one of the homologs is already known. The function of a gene can be determined experimentally by knocking out (deleting) or mutating that specific gene individually in a cell and observing the resulting phenotype.
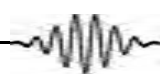
of human gene therapy is how genes might someday be used, modified, or even changed to correct various human diseases.

Genome Editing 'Gedanken' (Thought) Experiment

In the early 1990s, Srinivasan Chandrasegaran at the Bloomberg School of Public Health, Johns Hopkins Medical Institutions, conceived a genome editing gedankenexperiment, which is shown in *Figure* 1. The experiment involved the targeted replacement of a mutant gene (shown in red) with a normal gene (shown in green) in a human cell. In the converse experiment, one could replace a normal gene with a mutant allele, essentially knocking out the gene in cells. We then asked the following questions: (1) Do we have the molecular tools necessary to carry out such an experiment in a human cell? and (2) If we had the necessary tools, could we successfully perform the experiment without killing the cell?

In the early 1990s, the answer to both questions was a definite 'no'. Restriction enzymes that usually recognize 4–6 bp sites and cut DNA are not useful for this task since these sites would occur much too frequently in the human genome. Even sites for rare cutters that recognize >8 bp, would occur once every 65,536 bp in the human genome and cut too many times, killing the cell. Hence, this needed programmable enzymes that could recognize 16–18 bp sites to induce a unique targeted double-strand break (DSB) within the human genome.

Genome Editing

Scientists have long sought site-specific manipulation of plant and mammalian genomes, including the human genome. How does one achieve targeted genome editing, also known as gene editing or genome engineering, of plant and mammalian genomes? Cells use the universal process of homologous recombination (HR) to mediate site-specific recombination to maintain their genomic in-

Scientists have long sought site-specific manipulation of plant and mammalian genomes, including the human genome. How does one achieve targeted genome editing, also known as gene editing or genome engineering, of plant and mammalian genomes?

Figure 1. Genome editing gedankenexperiment (thought experiment). The objective of the experiment is the targeted replacement of a mutant gene (shown in red) with a normal gene (shown in green) in a human cell. In the converse experiment, one could replace a normal gene with a mutant allele, essentially knocking out that gene in cells. If the experiments are performed in germline cells, one can produce transgenic organisms.
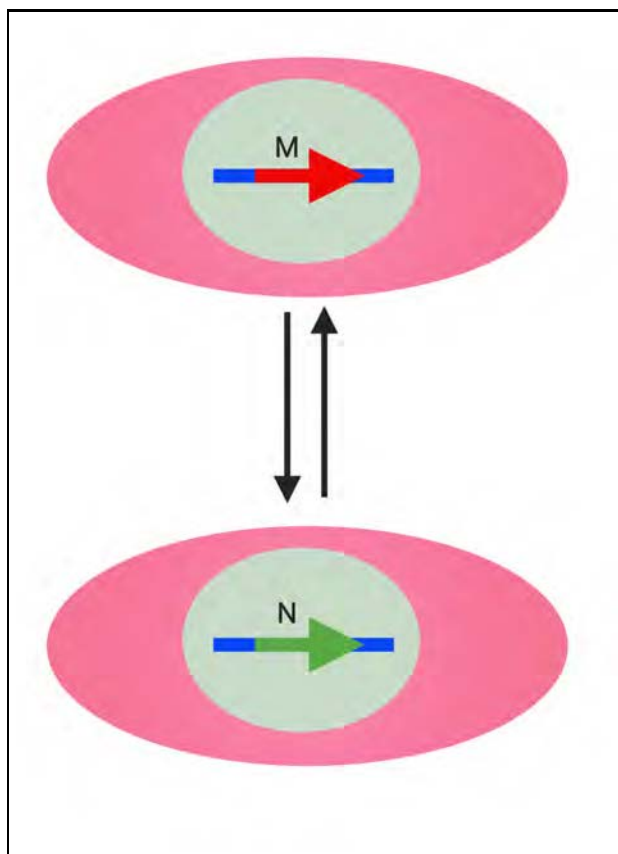
tegrity, especially during the repair of double-strand breaks (DSBs). DSBs are otherwise lethal to cells since they can scramble the encoded information of the genome. DSB repair of a damaged chromosome by HR in a cell is the most accurate form of repair, which works by copy and paste mechanism, usually using the homologous DNA segment from the undamaged sister chromatid as a template.

> Gene targeting—the process of replacing a gene by HR—uses an extrachromosomal fragment of donor DNA and invokes the cell's own repair machinery for gene conversion.

Gene targeting—the process of replacing a gene by HR—uses an extrachromosomal fragment of donor DNA and invokes the cell's own repair machinery for gene conversion. While spontaneous gene targeting is very efficient in bacteria and yeast cells, it is a very inefficient process in plant and mammalian cells; only one in a million treated cells undergo the desired gene modification. However, the introduction of a chromosomal DSB at a unique site

within a genome has been shown to stimulate HR by several orders of magnitude at that local site. The challenge then was to develop a general means of introducing a targeted DSB at a unique chromosomal locus of plant and mammalian genomes to induce homology-directed repair (HDR) at that site with the exogenous donor DNA provided by the investigator.

Towards this end, Chandrasegaran lab embarked on the quest to create programmable molecular scissors that were needed to perform the genome editing (aka gene editing or genome engineering) experiment in plant and mammalian cells. In 1996, Chandrasegaran lab reported the creation of the first programmable nucleases called zinc finger nucleases (ZFNs) [1, 2]. In 2001, Chandrasegaran lab, then in collaboration with Dana Carroll lab in Utah, showed that ZFNs by inducing a targeted DSB, stimulate local site-specific HR in frog oocytes, ushering the era of site-specific genome editing [3].

In 2021, it is safe to say that we have the necessary tools (including) ZFNs, TALENs and CRISPR-Cas9) available for targeted manipulation of the human genome; and that the gedankenexperiment has been performed successfully by many labs in many previously recalcitrant organisms that include plants, animals, insects and human cells [4]. Such genome editing experiments are now conducted routinely by labs around the globe.

Engineering Chimeric Nucleases for Genome Editing

Zinc Finger Nucleases (1996)

The first truly programmable were the zinc finger nucleases (ZFNs), which showed that arbitrary DNA sequences could be targeted for cleavage by protein engineering, ushering in the era of genome editing. ZFNs resulted from basic research on zinc finger proteins (ZFPs) and FokI restriction enzyme that revealed a bipartite structure with a separable DNA binding domain and a nonspecific cleavage domain. Chandrasegaran lab's studies on proteolytic fragments of FokI endonuclease revealed a 41 kDa N-

terminal DNA-binding domain and a 25 kDa C-terminal domain with non-specific DNA-cleavage activity. Crystal structures of the native FokI and FokI bound to DNA confirmed the modular nature of FokI endonuclease [5]. The structures revealed that the cleavage domain is sequestered in a piggyback fashion by the recognition domain, which was consistent with the DNA foot-printing analysis. Thus, FokI crystal structures were in complete agreement with the model derived from rigorous biochemical studies. Furthermore, the native FokI enzyme crystallized as a dimer, through the interaction between aspartic acid–arginine residues at the dimer interface, occurring between the cleavage domains [5].

The modular nature of FokI endonuclease suggested that it might be feasible to construct 'chimeric' nucleases with novel sequence-specificities by linking other DNA-binding proteins to the cleavage domain of FokI endonuclease. This indeed proved to be the case. Chandrasegaran lab reported the construction of the first chimeric nuclease by linking the *Drosophila Ubx* homeodomain to the cleavage domain of the FokI enzyme. Chandrasegaran lab then reported the creation of novel ZFNs by linking two different three zinc finger proteins to the FokI cleavage domain [1]. ZFNs are artificial proteins that combine FokI endonuclease activity with the ability of zinc finger motifs (ZFs) to specifically recognize a base triplet in DNA.

DNA Recognition by ZF Motifs

The Cys_2His_2 ZF motif can target specific sequences by virtue of its unique 30 amino acid residues with a $\beta\beta\alpha$ structure that is stabilized by a zinc ion (*Figure* 2a) [6]. Each ZF motif recognizes a 3- to 4-bp DNA sequence by inserting the α-helix into the major groove of the double helix (*Figure* 2b). Pavletich and Pabo reported the crystal structure of a set of 3-fingers from the mouse transcription factor Zif268, which gave a glimpse of ZFs recognize their cognate sequence (*Figure* 2c) [13]. Amino acids within the α-helix (positions -1, +1, +2, +3, +4, +5 and +6) of the ZF motif that makes sequence-specific contacts to DNA (*Fig-*

(a) Zinc Finger Motif (ZF)

(b) DNA Recognition by ZF Motif

Pabo et al., Annual Rev Biochem (2001)

Miller and Pabo, J Mol Biol (2001)

(c) Zinc Finger Protein (ZFP)

(d) Modeling ZFNs bound to DNA

Pabo et al., Annual Rev Biochem (2001)
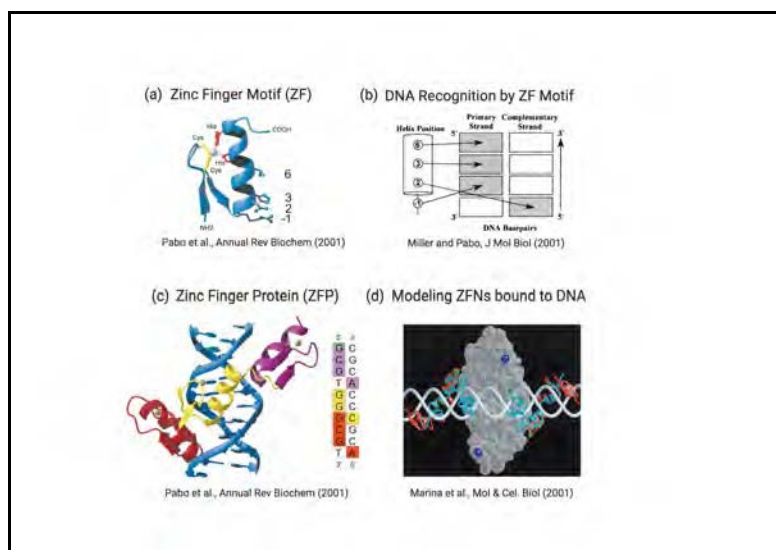
Marina et al., Mol & Cel. Biol (2001)

Figure 2. Zinc finger nucleases (ZFNs) combine the nonspecific cleavage domain of the FokI restriction enzyme with designed or selected ZFPs to provide a general mechanism to introduce a site-specific DSB within the genome. **(a)** The structure of zinc finger motif (ZF). **(b)** Amino acid residues within the ZF motif that make contact with DNA. **(c)** The structure of three ZF motifs that are fused in tandem to form a 3-finger ZFP and the cognate DNA sequence. **(d)** Modelling ZFPs fusion to FokI cleavage domain to form ZFNs. Dimerization of FokI nuclease domains is required to cut DNA. Two ZFNs bind to their cognate sites (inverted repeats) on DNA with dimerization of the FokI catalytic domains to cut the target site.

ure 2b), can be changed while maintaining the remaining amino acids as a consensus backbone to generate ZF motifs with new triplet sequence specificities. Most ZFs make contact with their target 3-bp site; however, when there is an aspartic acid present at the +2 position of the α-helix, it can reinforce an adenine or cytosine residue outside the 3-bp site at the next base of the non-contact strand of DNA via cross-strand contact, changing ZFs recognition to a 4-bp site (*Figure* 2b & 2c). This ZF contact outside the 3-bp site further influences the specificity of neighbouring ZFs. This complicates the generation of ZFPs by simple modular design, wherein each ZF recognizes a triplet sequence. In theory, three ZF motifs that act independently of each other can be linked together in tandem to generate a ZFP that binds to a 9-bp site, which is a composite of the individual DNA triplet subsites recognized by each of the three ZF motifs. However, in practice, since some ZF motifs influence their neighbours through the cross-strand contact, the design and selection of each ZF have to be performed in a context-dependent fashion to obtain highly sequence-specific ZFPs, which is laborious, time-consuming and makes their production very expensive.

ZFNs

ZFNs combine the nonspecific cleavage domain of the FokI restriction enzyme with the designed or selected ZFPs to provide a general mechanism to introduce a site-specific DSB into the genome (*Figure* 2d) [2–4]. Furthermore, binding of two 3-finger ZFN monomers, each recognizing a 9-bp inverted site, is necessary because dimerization of the FokI cleavage domain is required to produce a DSB [2–4]. Thus, three-finger ZFNs effectively have an 18-bp recognition site, which is long enough to specify a unique address within the human genome. Two ZFNs with different sequence specificities can bind to appropriately positioned heterologous sites in tandem and induce a targeted DSB within the human genome (*Figure* 3a) [2–4]. The sequence specificity could be enhanced further by increasing the number of ZF motifs within the ZFPs, and hence ZFNs (as shown in *Figure* 3a).

Transcription Activator-like Effector Nucleases (2010)

The accidental discovery of a novel DNA-binding module, called TALE (Transcription Activator-Like Effector) modules found in the plant virus from *Xanthomonas* bacteria, led to the construction of TALENS (*Figure* 3b) [7–9]. The TALE central repeat domain consists of repeating units of 30–33 amino acids. The TALE modules are largely identical except for highly variable amino acids at positions 12 & 13, known as the repeat-variable diresidues (RVDs). Unlike ZF motifs which recognize 3- or 4-bp DNA sites, each TALE module recognizes a single nucleotide [7, 8]. The recognition specificity of the TALE modules is determined by the RVDs: NI recognizes adenine, HD recognizes cytosine, NG or HG recognizes thymine, and NN recognizes guanine or adenine (*Figure* 3b). Unlike the ZFs, the recognition of DNA by TALEs is largely independent of the neighbouring modules. The DNA recognition code thus provides a simple one to one correspondence between an array of TALE units and the nucleotide sequence of the cognate site. This simple recognition code and modular nature of TALE motifs make them a versatile
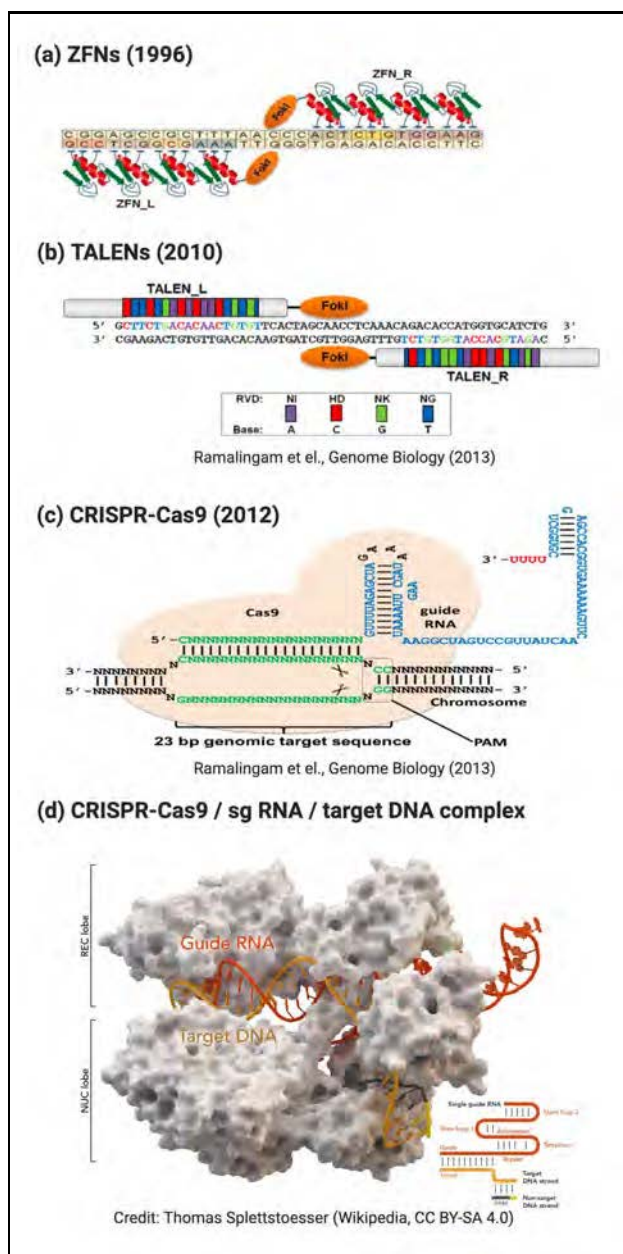
(a) ZFNs (1996)

(b) TALENs (2010)

RVD: NI HD NK NG
Base: A C G T

Ramalingam et el., Genome Biology (2013)

(c) CRISPR-Cas9 (2012)

23 bp genomic target sequence — PAM

Ramalingam et el., Genome Biology (2013)

(d) CRISPR-Cas9 / sg RNA / target DNA complex

Credit: Thomas Splettstoesser (Wikipedia, CC BY-SA 4.0)

Figure 3. Gene editing tools include ZFNs, TAL-ENs and CRISPR-Cas9. ZFNs were the truly targetable reagent that was developed first; they laid the foundation for targeted genome editing of various organisms. ZFNs and TALENS require protein engineering for target recognition, which is difficult by nature. CRISPR-Cas9 target recognition is done by RNA, which is much easier to design. **(a)** ZFNs binding to their cognate sites.**(b)** TAL-ENs binding to their cognate sites. **(c)** Schematic of CRISPR-Cas9 and sgRNA binding to their target site. The crystal structure of Cas9/tracrRNA/crRNA complex bound to the target DNA is also shown. The year in which the gene-editing tools were introduced is shown in brackets.

tool for generating custom nucleases.

TALENs are designed and constructed in a modular style very similar to that of ZFNs, using the immense experience gained from ZFN architecture (*Figure* 3a & 3b). While ZFNs use ZF

motifs as DNA binding modules, TALENs use TALE motifs as DNA-binding modules, but both use the FokI catalytic domain as the DNA cleavage module [4]. TALENs and ZFNs appear to have similar cleavage efficiencies. The genes encoding TALENS are about three times larger than ZFNs; this is because TALEs are similar in size to ZFs but recognize a single base, while ZFs recognize 3–4 bp sequences. Since TALENS are easier to generate than ZFNs, they are also relatively inexpensive.

RNA-guided CRISPR-Cas9 (2012)

Discovery of CRISPR Repeat Sequences

In 1987, Japanese researchers studying the isozyme alkaline phosphatase gene observed the presence of five highly homologous sequences of 29 nucleotides arranged as direct repeats with 32 nucleotides spacing downstream of the gene.

In 1987, Japanese researchers studying the isozyme alkaline phosphatase gene observed the presence of five highly homologous sequences of 29 nucleotides arranged as direct repeats with 32 nucleotides spacing downstream of the gene. The first repeat was included in the putative transcription termination site and showed less homology to the other four. The well-conserved sequences exhibited a 14 nucleotides dyad symmetry, which were similar to the ones that were previously found in *E. coli* and *S. typhimurium*, but the significance of the repeats was unknown. In 1992, Francisco Mojica doing his thesis work at the University of Alicante was studying microorganisms belonging to the Archaea family, halophile microorganisms that require high-salt conditions to survive. Mojica and his colleagues were interested in understanding how Archaea can grow in high salinity and how they adapt to changes in salinity. DNA sequencing led to the observation that the halophiles' genome possessed a series of regularly spaced repeats, which they labelled as tandem repeats (TREPS) that were later renamed as CRISPR (an acronym for Clustered Regularly Interspaced Short Palindromic Repeats). They also observed that these repeated regions in the halophiles were actively transcribed indicating that the cell was reading this information in each of the growing conditions that they tested and that the repeats were important for the cell. Furthermore, when sequencing one strain of *E. coli*, Mojica discovered that there were sequences between the

repeats known as 'the spacer regions of CRISPRs that matched the sequence of a particular virus. These DNA sequences protected the bacteria from being infected by viruses carrying the same sequence in its genome. The virus could not infect the cell, which suggested that it was an adaptive immune system evolved by prokaryotes. Exploration of sequencing data revealed that this was the case in many other organisms as well. Further research by scientists led to the identification of Cas enzyme systems that were involved in the acquisition of spacer sequences and the mechanism of adaptive immunity in prokaryotes.

Harnessing CRISPR-Cas9 for Genome Editing

Bacteria and archaea thus have evolved an adaptive defence mechanism that use CRISPR/Cas systems to degrade complementary sequences present within invading virus and plasmid DNAs [10]. Type II CRISPR/Cas systems rely on the integration of foreign DNA fragments into the CRISPR loci, which upon transcription and processing result in CRISPR RNAs (crRNA), which then anneal to trans-activating crRNAs (tracrRNAs) enabling Cas proteins to direct sequence-specific degradation of foreign DNA (*Figure* 3c). In 2012, scientists established that the Cas9-crRNA complex functions as an RNA-guided endonuclease, with RNA-directed target DNA recognition and protein-mediated DNA cleavage, paving an easy way to engineer universal programmable RNA-guided endonucleases [11]. Virginijus Šikšnys (at Vilnius University, Lithuania) and co-workers reported that the Cas9-crRNA complex functions as an RNA-guided endonuclease with RNA-directed target sequence recognition, and protein-mediated DNA cleavage, paving the way for the engineering of universal programmable RNA-guided DNA endonucleases [12]. Jennifer Doudna, Emmanuelle Charpentier and co-workers at the University of Berkeley showed that Cas9-mediated cleavage can also function efficiently using the fusion of crRNA and tracrRNA to form a single guide RNA (sgRNA), which greatly simplified the application of the CRISPR-Cas9 system for genome editing (*Figure* 3c) [12]. Subsequently, Feng Zhang lab in the Broad Institute at MIT,

Bacteria and archaea have evolved an adaptive defence mechanism that use CRISPR/Cas systems to degrade complementary sequences present within invading virus and plasmid DNAs.

along with several other labs, reported that they could engineer the CRISPR-Cas9 system to function with custom sgRNA in human cells to direct sequence-specific cleavage [13]. The targeting efficiency of the endogenous loci in human cells was better than that observed with ZFNs and TALENS targeting the same loci. Furthermore, genome editing of multiple sites within the human genome was possible by simultaneous introduction of several sgRNAs into the cell.

Unlike ZFNs and TALENS, which require protein engineering of two nucleases, CRISPR-Cas9 depends on RNA-DNA recognition for genome editing. RNA design is much simpler and easier when compared to protein engineering of two nucleases, each recognizing one-half of the genomic target sequence as in the case of ZFNs and TALENS. The advantages of the CRISPR-Cas9 system include the ease of RNA design for new targets, the dependence of a single, constant Cas9 protein and its ability to address multiple targets simultaneously with multiple guide RNAs. These have led to its wide adoption in research laboratories around the world. The CRISPR-Cas9 methodology is also very inexpensive, making it very affordable for small labs.

The 2020 Nobel Prize for Chemistry was awarded to Jennifer Doudna and Emmanuelle Charpentier[1] for harnessing the CRISPR-Cas9 system for genome editing and providing a simplified technique. Many scientists were surprised that ZFNs' original contribution that includes launching the era of genome editing, and laying the foundation for genome editing, was not recognized to share the 2020 Nobel Prize.

Development of CRISPR-Cas Diagnostic Kits

Rapid detection of nucleic acids is an essential part of clinical diagnostics. CRISPR-based diagnostic applications have been reported by several labs. Two such diagnostic systems are described briefly here. Feng Zhang lab in Broad Institute at MIT, USA, reported a diagnostic platform called SHERLOCK (an acronym for Specific High Sensitivity Enzymatic Reporter unLOCKing) that

combines nucleic acid pre-amplification with CRISPR-Cas9 for specific amplification of DNA and RNA sequences. The method comprises of two steps: (1) recombinase-mediated polymerase pre-amplification of DNA or RNA sequences that is followed by (2) Cas13- or Cas12-mediated detection using fluorescent and calorimetric readouts that provide results in less than an hour. Maiti, Chakraborty and coworkers at CSIR-IGIB, New Delhi, India, have reported a diagnostic platform called FELUDA (an acronym for FnCas9 Editor Linked Uniform Detection Assay) to detect coronavirus during the COVID-19 outbreak. This system also pre-amplify the RNA extracted from the nasal swab of patients using a single-step RT-PCR. This approach then uses Fn-Cas9 that recognizes a different PAM (Protospacer Adjacent Motif) sequence than SpCas9. The FELUDA mix is prepared by incubating the dead FnCas9 protein (in which the active sites of the enzyme are mutated), guide RNA that is specific for the coronavirus target, and the amplified viral DNA. The dipstick contains gold nanoparticles that bind to the FELUDA complex. The gold particle-FELUDA complexes are then captured by streptavidin to constitute the test line on the dipstick. The unbound particles are captured on the control line. If colour develops only on the control line of the dipstick, it indicates the patient is negative. If colour develops on both control and test lines, it indicates that the patient is positive for coronavirus infection. The dipstick test takes 1–2 minutes to perform. Several labs have developed other diagnostic applications, which are not discussed here.

Off-target effects of programmable nucleases is a continued concern in genome editing.

Off-target Effects of Programmable Nucleases

Off-target effects of programmable nucleases is a continued concern in genome editing. Early ZFN studies suggested that some ZFNs could bind and form dimers at sequences other than those for which they were designed, and cleavage at secondary targets was the cause of the cell lethality. Off-target cleavage has been confirmed in many other situations with TALENs and CRISPR-Cas9 as well by analyzing the genomic sequences for induced mutations. Strategies have been designed to enhance the speci-

ficity of all the programmable nucleases. For ZFNs and TALENs, mutations were introduced to destabilize the dimer interface of the FokI cleavage domains that allow heterodimers to form, but not homodimers. This significantly reduced the toxicity of ZFNs and TALENS and was quite effective in reducing the off-target effects. With CRISPR-Cas9, truncating sgRNA had a positive effect on specificity, apparently because a small number of mismatches in a shorter duplex has a larger destabilizing effect at secondary sites [4]. Other approaches include introducing a mutation at one of the nuclease active sites of Cas9 converting into a nicking enzyme. Providing two sgRNAs directed to targets in close proximity to the DNA leads to paired nicking, which is quite effective in inducing NHEJ and HDR while increasing specificity by demanding simultaneous recognition by both guide RNAs [4]. How much off-target cleavage matters depends on the situation in which the programmable nucleases are used. For many research applications, the designed alteration in the target is most important, accommodation can be made for things happening in the background. To attribute a phenotype to an engineered change, however, one must use independently derived mutant alleles or rescue the phenotype by complementation [4]. For applications to food organisms and human therapeutics, much more stringent criteria must be applied. When cloned cells or whole organisms are edited, whole-genome sequencing can reveal all induced sequence changes [4]. Whole-genome sequencing of mixed cell populations needs to be very deep in order to reveal infrequent but potentially harmful changes. Several methods developed recently can also capture cleavage sites more broadly, and they are quite powerful and revealing [4].

Applications of Genome Editing to Life Sciences

Genome editing techniques—ZFNs, TALENs and CRISPR-Cas9—all deliver a targeted chromosomal DSB within cells to stimulate recombination (*Figure* 4). Cells have evolved two classes of mechanisms to repair DSBs: (1) non-homologous end joining (NHEJ); and (2) homology-directed repair (HDR). NHEJ is an
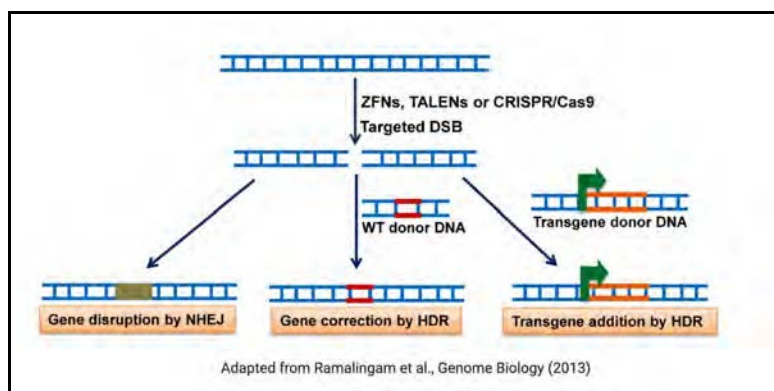
Figure 4. Genome engineering using ZFNs, TALENS and CRISPR-Cas9. Targeted genome editing by non-homologous end joining (NHEJ) or by homologous recombination (HR) using gene editing tools. In homology-directed repair (HDR) experiment, cells are transfected with both gene editing tool and a wild-type DNA segment (for mutation correction) or transgene donor DNA for targeted addition. For targeted knock-out of a gene, cells are transfected with gene editing tool alone.

error-prone repair mechanism that often results in variable lengths of insertion and deletion mutants (also known as 'indels'), generating a pool of mutants. Thus, NHEJ can be used to knock out genes. Nuclease-based breaks are efficiently mutated during NHEJ because perfectly rejoined sequences will be cut again until they acquire an indel, after which they no longer can be cleaved. HDR normally relies on recombination with homologous sequences of the undamaged sister chromatid, but cells can be tricked into using the investigator-provided donor DNA. HDR leads to the introduction of precise alteration to the genome that is specified by the donor DNA template.

Genome-editing can be used to correct, introduce or delete almost any DNA sequence in many different types of cells and organisms (*Figure* 4). Genome-editing (also called gene-editing) seeks to modify the genes of living organisms to understand gene function and develop ways to use genes to treat genetic or acquired diseases.

A few examples of how genome-editing technologies are revolutionizing life sciences are discussed below.

Anti-browning Mushroom

Scientists at the University of Pennsylvania reported engineering of the edible fungus—the common white button mushrooms— to resist browning (*Figure* 5a). It was achieved by knocking out

Genome-editing can be used to correct, introduce or delete almost any DNA sequence in many different types of cells and organisms. Genome-editing (also called gene-editing) seeks to modify the genes of living organisms to understand gene function and develop ways to use genes to treat genetic or acquired diseases.

one of the six families of genes that encode polyphenol oxidase (PPO) enzymes. The targeted PPO gene was inactivated by deleting a few base pairs in the gene of the mushroom's genome using CRISPR-Cas9; this resulted in the reduction of the PPO enzyme's activity by 30% in the anti-browning mushrooms.
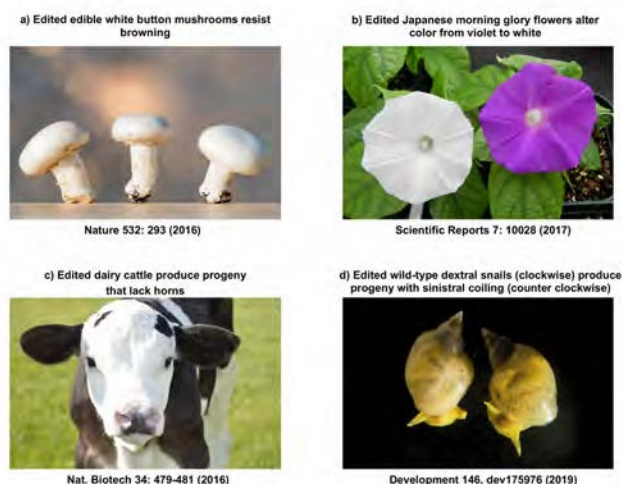
In 2016, the US Department of Agriculture (USDA) decided that it will not regulate a mushroom that has been modified genetically using CRISPR, making it the first CRISPR-edited organism to get approval from the US government. The decision means that the mushroom can be cultivated and sold without passing through the agency's regulatory process. Prior to this, several other plants, made using ZFNs and TALENs, have passed the USDA regulatory process. The gene-edited mushroom did not trigger USDA oversight because it does not contain any foreign DNA, such as virus or bacteria. Such organisms were necessary for genetically modifying plants and fungi in the 1980s and 1990s when the US government developed the framework for regulating GMOs.

In 2016, the US Department of Agriculture (USDA) decided that it will not regulate a mushroom that has been modified genetically using CRISPR, making it the first CRISPR-edited organism to get approval from the US government.

Changing Flower Colour

Japanese scientists have used CRISPR-Cas9 to change the flower colour of the ornamental plant, Japanese morning glory (*Ipomoea nil* or *Pharbitis nil*) from violet to white (*Figure* 5b). Japanese morning glory or Asagao was used because its genome is sequenced, and DNA transfer methods are well established to carry out gene editing. The researchers targeted a single gene, *dihydroflavonol-4-reductase-B* (*DFR-B*), encoding an anthocyanin biosynthesis enzyme, which is responsible for the colour of the plant's stems, leaves and flowers. Two other closely related genes (*DFR-A* and *DRF-C*) were adjacent to *DFR-B*. CRISPR was used to specifically and accurately target and inactivate enzyme *DFR-B* gene without altering *DFR-A* and *DRF-C*. Disruption of *DFR-B* should de-activate the enzyme, resulting in the absence of the colour pigment anthocyanin. The CRISPR-Cas9 system was inserted into tissue-cultured embryos of Japanese morning glory plants using the plant bacterium *Rhizobium*. The *DFR-B* enzyme was successfully inactivated in 75% of the transgenic plants with

Japanese scientists have used CRISPR-Cas9 to change the flower colour of the ornamental plant, Japanese morning glory (Ipomoea nil or Pharbitis nil) from violet to white.

Figure 5. A few selected examples of genome editing applications in life sciences. **(a)** The edible white button mushrooms have been edited to resist browning by knocking out one of the PPO genes. **(b)** The ornamental Japanese morning glory flower colour has been edited from violet to white by knocking out the *DHR-B* gene. **(c)** Scientists have edited dairy cattle to lack horns by disrupting the POLL locus of the cow genome, which is responsible for horn development. **(d)** Editing the *Ls-dia1* gene in natural dextral (clockwise) pond snails led to off-springs with the rare sinistral (anticlockwise) shells, suggesting that *Ls-dia1* is key to determining the creatures' 'handedness'.

green stems and white flowers. Non-transformed plants with an active enzyme had violet stems and flowers. Furthermore, genetic analyses confirmed that the *DFR-B* target gene had been altered in the transgenic plants, with either DNA insertions or deletions in both copies of the genes (bi-allelic mutants). *DFR-A* and *DFR-C* genes were unaltered with no mutations.

Hornless Dairy Cattle

Over the last 10,000 years, humans selectively bred cattle to produce more meat or milk, selecting distinct genetic variations over non-domesticated cattle. Cattle have evolved horns to defend themselves against predators and gain a reproductive advantage over competing peers by signifying health and vitality to females. Horns on dairy cattle can injure their handlers and other cattle. Physical dehorning of cattle is widely practised, which is a bloody and painful process that pits animal rights activists against it. Conventional selective breeding is a slow and expensive process, usually taking many generations before the trait is at a high frequency in the population. However, some cattle breeds, like the Angus beef breed, are hornless due to a natural genetic variation or allele that arose in the distant past. Scientists have shown

Dairy cattle do not carry the *Polled* allele, and hence these animals have horns. Genome-editing has offered a simple and efficient way to transfer the small and natural *Polled* allele from beef cattle to dairy cattle, and quickly produce dairy cattle without horns and importantly while preserving the elite dairy cattle genetics such as high yielding milk cows that were generated by classical breeding.

that the mutant allele, called *Polled*, prevents horn growth when an individual inherits one or two copies of the *Polled* gene from its parents, i.e., the *Polled* allele is dominant to the usual *Horn* allele, at one specific chromosomal locus of the cow genome. The complete DNA sequence of the cow genome, including its genes are known. Dairy cattle do not carry the *Polled* allele, and hence these animals have horns. Genome-editing has offered a simple and efficient way to transfer the small and natural *Polled* allele from beef cattle to dairy cattle, and quickly produce dairy cattle without horns and importantly while preserving the elite dairy cattle genetics such as high yielding milk cows that were generated by classical breeding. As an alternative, gene-editing technologies can precisely and permanently eliminate horns from dairy cattle, without sacrificing the elite dairy production genetics. Thus, genome editing offered a genetic alternative to removing the horns physically. Using TALENs, the researchers demonstrated the birth of two hornless male cattle in 2015, in reproductive crosses where the offspring normally should be horned (*Figure* 5c). These bulls then sired six hornless calves born in 2017. Thus, genome-editing technology is a safe and relatively fast way of transferring a small and naturally occurring genetic variant from one breed of cattle into another. Gene-editing can speed up what humans have been doing for thousands of years using the very slow process of selective breeding for desirable traits.

The greatest challenge, however, to using genetically edited hornless cattle in the dairy industry is the approval of the Food and Drug Administration (FDA), which classifies these cattle in the same category as a new animal 'drug'. It must be pointed out that Angus beef breed of cattle has been consumed by humans for a thousand years, and the Polled genetic variation was transferred into a breed of dairy cattle, whose milk has been consumed by humans for hundreds of years.
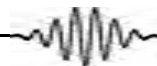
Altering Snail Shell Spiral

Japanese scientists have used CRISPR–Cas9 gene-editing to knock out a gene called *Lsdia1* in the snail, *Lymnaea stagnalis*. *Lsdia1* encodes a protein that helps to govern a cell's internal framework. Snails normally have a dextral shell that spirals out from the centre in a clockwise direction (right-handed). But a few left-handed snails have sinistral shells that coil in the opposite counterclockwise direction. CRISPR genome-editing has helped to pinpoint *Lsdia1* as the gene that is involved in establishing an embryo's body plan.

The researchers found that snails missing the gene had offspring with the rare anticlockwise shells, suggesting that *Lsdia1* is key to determining the 'handedness' (*Figure* 5d). These snails passed on the mutated gene and shell-type to subsequent generations. Those with *Lsdia1* functioning copies of the gene had shells that coiled the normal clockwise direction. *Lsdia1* gene appears to influence the pattern of cell division from the stage at which snail embryos divided from one cell into two. Scientists believe that research on *Lsdia1* could shed light on a rare human condition called *situs inversus*, in which the positions of the organs in the body are reversed.

Applications of Genome Editing to Human Therapeutics

Genome-editing approaches when coupled to stem-cell technology, can have a huge impact on the future of personalized precision medicine. A combination of these approaches has been tested in clinics for their therapeutic benefit for some diseases, using either ex vivo or in vivo gene therapy. In ex vivo therapy, the defective gene is repaired in vitro in patient-derived stem cells and then progenitor cells are infused back into the patient. In the case of in vivo therapy, the genome-editing reagents are packaged in suitable delivery vehicles such as viral vectors or nanoparticles and then delivered directly into the patient to achieve in vivo editing of the targeted gene in the patient's cells within the desired tissues. Below, we outline a few selected biomedical applications

> Genome-editing approaches when coupled to stem-cell technology, can have a huge impact on the future of personalized precision medicine. A combination of these approaches has been tested in clinics for their therapeutic benefit for some diseases, using either ex vivo or in vivo gene therapy.

of genome editing for monogenic diseases in humans and look at the future prospects in the context of cellular therapy and regenerative medicine using blood disorders as examples. ZFNs are the most advanced in the clinic to treat several human conditions.

In the USA, the National Institutes of Health (NIH) supports gene-editing approaches in somatic cells for many diseases. Since somatic cells are not involved in human reproduction, genomic changes in them will not be inherited by future generations. Genome-editing approaches are funded and being pursued as part of NIH's Cure Sickle Cell Initiative.

Blood Disorders

Blood disorders encompass a range of diseases, many of them caused due to genetic mutations. Presently, allogenic bone marrow transplantation (BMT) is the only curative therapy available for a range of blood disorders. Over the last several years, the number of treated patients using BMT has significantly increased. Though numerous blood disorders can be cured, the treatments have the potential for significant complications. An emerging alternative to allogeneic BMT is the precise modification of hematopoietic stem/progenitor cells (HSPCs) using the targeted genome editing approaches.

β-hemoglobin Disorders

β-hemoglobin disorders are a group of autosomal recessive blood disorders caused by genetic mutations in the adult haemoglobin (*HBB*) gene on chromosome 11. *HBB* gene encodes the β-globin protein, which is a vital component in the functional haemoglobin present inside red blood cells (RBCs) and is essential for oxygen delivery to the tissues. Mutations in the *HBB* gene result in deficiency of functional RBCs (anaemia), multi-organ dysfunction and death. β-thalassemia (BT) and Sickle Cell Disease (SCD) are the most prevalent disorders in β-hemoglobinopathies worldwide. SCD results from the production of dysfunctional haemoglobin (HbS) protein, while β-thalassemia is the result of insufficient

production of functional HBB protein. There are two different methods to develop genome-engineering based therapeutic strategies for β-hemoglobin disorders, which are as follows: (1) Genetic correction of the adult *HBB* gene and (2) Enhancing fetal haemoglobin (HbF) production in patient-derived hematopoietic stem and progenitor cells (HSPCs).

Genetic Correction of HBB *Gene*

As the *HBB* gene possesses two major disease-causing mutations, genetic correction of this gene is a primary focus of the genome editing based therapies. In this approach, sickle cell mutation in patient-derived HSPCs is corrected ex vivo by donor DNA-based HDR, employing gene-editing nucleases. The corrected HSCs are expanded and then transplanted back to the same patients (*Figure* 6). This autologous transplantation of genetically corrected HSCs eliminates the need for allogeneic transplantation. Hoban et al., reported the first successful correction of SCD mutation in patient-derived HSPCs using ZFNs [14]. In this study, the corrected HSPCs maintained their ability to engraft non-obese diabetic (NOD)/SCID/gamma mice (NSG mice) and showed multi-lineage differentiation potential. Subsequently, Dever et al., demonstrated efficient correction of SCD mutation in patient-derived HSPCs using CRISPR-Cas9, sgRNA ribonucleoprotein, and rAAV6 for HR donor delivery [15]. The corrected HSPCs successfully differentiated into erythrocytes expressing normal *HBB* levels. The above gene-editing studies in HSPCs have performed on the use of viral vector donors for HDR. However, DeWitt et al., demonstrated the selection-free *HBB* mutation correction in SCD patient-derived HSPCs using optimized design and delivery of ribonucleoprotein (RNP) complex consisting of SpCas9 and sgRNA, along with single-stranded DNA oligonucleotide. In addition to HSPC, several research groups have shown that gene editing nucleases can be used to stimulate effective *HBB* gene correction in SCD patient-derived induced pluripotent stem cells (iPSCs).

BT is caused by the array of mutations in the adult *HBB* gene,

Figure 6. A schematic diagram showing the ex vivo genome-editing based autologous cell replacement therapy for blood disorders. CD34+ HSPCs are isolated from a patient, and the defective gene is then repaired, or a beneficial mutation is generated using gene editing nucleases through viral or non-viral approaches. Edited cells are characterized in vitro and then infused back into the bone marrow of the patient.
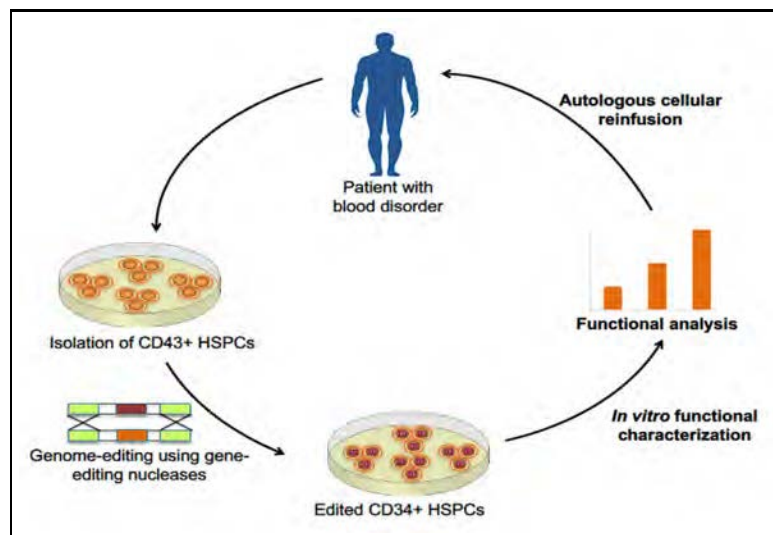
which makes to develop a mutation-specific correction for each and every BT mutation. The situation gets worse in the case of patients having two compounds for each and every BT mutation. The situation gets worse in the case of patients having two compound teterozygous mutations. Xu et al., has successfully targeted IVS-110 G>A and IVS2-654c>T using CRISPR-Cas9 and CRISPR Cas9/Cpf1 respectively in BT patient's HSPCs. To date, no attempt has been made to correct the BT mutation in patient-derived HSPCs using HDR-based genetic correction.

Enhancing Fetal Hemoglobin Production in Adults HSPCs

The production of γ-globin subunit leads to the synthesis of fetal haemoglobin (HbF) in early life (fetal stage), but is silenced and replaced by the production of β-globin resulting in the synthesis of adult haemoglobin (HbA). Usually, γ-globin production is very minimal (<1%) in adults, but there are some healthy individuals with higher HbF levels, and this condition is known as hereditary persistence of fetal haemoglobin (*HPFH*). Remarkably, observations suggest that the patients of SCD and BT show mitigation in clinical manifestation if there is coinheritance with heterozygous for *HPFH*, which occurs due to certain mutations in the beta-globin cluster at chromosome 11 that lead to an enhanced ex-

pression of the γ-globin genes namely *HBG1* and *HBG2*. These mutations disrupt either expression of γ-globin repressor genes or their binding sites and thereby help reactivate HbF. Most of the reported *HPFH* mutations have been classified into two subtypes: (1) Non-deletional *HPFH* mutations are single nucleotide polymorphisms (SNPs) found upstream of *HBG1/2* genes. 2) Deletional *HPFH* mutations that are the deletions of a wide range (13–106 kb) DNA segment, mostly found in the β-globin gene cluster on chromosome 11, encompassing γ-globin repressor binding sites and competitive genes (*HBD* and *HBB*). Hence generating *HPFH* or HPFH like genotype to reactive HbF in patient-derived HSPCs using genome-editing, has enormous promise for curative treatment for β-hemoglobinopathies.

Hemophilia

Haemophilia is an X-linked monogenic disorder caused due to deficiency in coagulation factors in the intrinsic coagulation cascade. Haemophilia is broadly classified into two types, namely— Haemophilia A and Haemophilia B, based on the site of genetic mutation. Haemophilia A is a more prevalent form of haemophilia, and occurs in 1 in 5000 live male births due to a mutation in the gene coding for factor VIII (FVIII), resulting in the loss of functional FVIII protein. Due to the high treatment costs, frequencies of infusions, and potential health burden, there is a need for alternative, more cost-effective therapies for haemophilia.

More studies are employing gene-editing nucleases for the correction of haemophilic mutations. One of the first studies that showed the possibility of genome-editing as a viable therapeutic strategy to treat haemophilia was by Li et al., 2011. They showed that ZFNs can induce DSBs efficiently when delivered directly to the mouse liver and that when co-delivered with an appropriately designed gene-targeting vector, they can stimulate gene replacement through both homology-directed repair and homology-independent targeted gene insertion at the ZFN-specified locus. They have demonstrated that the level of gene targeting achieved was sufficient to correct the prolonged clotting times in a mouse

Haemophilia is an X-linked monogenic disorder caused due to deficiency in coagulation factors in the intrinsic coagulation cascade. Studies employ gene-editing nucleases for the correction of haemophilic mutations. One of the first studies that showed the possibility of genome-editing as a viable therapeutic strategy to treat haemophilia was by Li et al.

model of haemophilia B and remained persistent after induced liver regeneration.

Human Immunodeficiency Virus

In 2014, the first clinical application of genome-editing involved the use of ZFNs to make human cells resistant to HIV-1 by disrupting *CCR5*, the co-receptor needed for the viral entry into cells. The gene-editing approach as a form of gene therapy for HIV was enabled by the finding that a small subset of population across the world, naturally has a 32 bp deletion in the *CCR5* gene (*CCR5Δ32*). This creates a premature stop codon, thereby resulting in truncated, non-functional CCR5 protein. Studies have shown that individuals with *CCR5Δ32* appear to be permanently resistant to HIV infection. In 2008, Perez et al., established the first proof-of-concept for *CCR5* gene-editing by delivering adenoviral vectors carrying both ZFNs into primary human CD+ T-cells. In 2010, Holt et al., then showed *CCR5* gene-editing in human CD34+HSPCs. Holt et al., also showed that *CCR5* depleted mice develop 0 patients with the mean modification efficiency of 17% in CD34+ HSPCs. Furthermore, mouse studies showed that edited HSPCs, when transplanted, demonstrate a selective resistance to HIV infection.

> In 2014, the first clinical application of genome-editing involved the use of ZFNs to make human cells resistant to HIV-1 by disrupting *CCR5*, the co-receptor needed for the viral entry into cells. The gene-editing approach as a form of gene therapy for HIV was enabled by the finding that a small subset of population across the world, naturally has a 32 bp deletion in the *CCR5* gene (*CCR5Δ32*).

Hunter's Syndrome (MPS II)

In 2017, Sangamo Therapeutics Inc. in the USA launched a Phase I clinical trial testing of ZFNs to correct Hunter's syndrome (MPS II) caused by iduronate-2-sulfatase enzyme deficiency, which is responsible for breaking down complex sugars in the body. Hunter's syndrome can cause abnormalities in the skeleton, heart and respiratory systems. The clinical trial was the first in vivo genome editing approach administered directly to patients, and the treatment was effective without side effects.

Chimeric Antigen Receptor (CAR) T-cell Therapy

TALENs have also been used in T-cell immunotherapy approaches to create 'off-the-shelf' universal donor T-cells that don't have to be developed for every cancer patient. One approach is the development of CAR T-cell therapies that use immune cells collected not from patients, but from healthy donors. The idea is to create so-called off-the-shelf CAR T-cell therapies that are immediately available for use and do not have to be manufactured for each patient. Cellectis (a French company) has launched a Phase I trial of its off-the-shelf CD19-targeted CAR T-cell product in the USA for patients with advanced acute myeloid leukaemia. The company's product has already been tested in Europe previously, including in two infants with ALL who had exhausted all other options. In both cases, the treatment was effective.

TALENs have also been used in T-cell immunotherapy approaches to create 'off-the-shelf' universal donor T-cells that don't have to be developed for every cancer patient.

Safety, Ethical and Policy Concerns

Applications of genome-editing have raised scientific, safety, ethical and policy concerns. The National Institute of Health (NIH) in the USA does not fund any use of genome-editing in human embryos. The US Congress has prohibited the use of federal funds for the creation of human embryos for research purposes or for research in which human embryos are destroyed. Furthermore, off-target effects of the genome-editing approaches pose serious and unquantifiable safety and ethical issues when altering the germline, which will affect future generations without consent. Also, a lack of compelling medical applications does not justify the use of genome-editing in the human embryo when other alternatives such as prenatal screening are available. The National Academies of Science, Engineering, and Medicine (NASEM) has issued a report, entitled *Human Genome Editing: Science, Ethics, and Governance* (2017), which recommends that clinical trials using gene-editing in embryos should be permitted only within a robust and effective regulatory framework, and only when certain stringent criteria are met.

The safety and bioethical concerns raised by use of genome-editing

in somatic gene therapy are same as that were raised for recombinant DNA technology and human gene therapy. These concerns are well-addressed by the current oversight structures in the United States. The US Food and Drug Administration's (FDA) regulatory authority oversight will now apply to clinical research involving genome-editing somatic cells in humans. Biosafety guidance for conducting genome editing research is provided by the *NIH Guidelines for Research Involving Recombinant or Synthetic Nucleic Acid Molecules.*

A draft document on *Genome Edited Organisms: Regulatory Framework and Guidelines for Risk Assessment* has been prepared by Dr S. R. Rao from the Department of Biotechnology (DBT), Ministry of Science and Technology, India. The proposed regulatory framework and guidelines for risk assessment is yet to be approved by the Indian Government. The Indian statutory committees for authorization and oversight of genome-editing research and products include IBSC, RCGM, GEAC and statutory market.

Summary

Genome-editing technologies have revolutionized life sciences. Their impact on biology, agriculture, ecology and personalized medicine have been tremendous and will be felt for years to come. Nobody thought that rewriting life was possible before. Genome-editing and synthetic biology have changed all that. These fields have enabled scientists to do precise genome surgery and create new organisms from scratch, respectively. It is a wonderful time for young undergraduate students, graduate students and medical professionals to enter these fields seeking a research career in biology and/or personalized precision medicine.

Acknowledgement

lore, India. He was affiliated to Professor D. N. Rao's lab in the Department of Biochemistry.

Box 1. Glossary

Conservative changes: Substitution of one amino acid by another, possessing similar biological properties.

Degenerate codons: Multiple codons that specify the same amino acid.

Genotype: Genetic information present within a cell of an organism. It contributes to the physical attributes or phenotype of the organism.

Heterozygous: A gene having two different types of alleles for a specific trait.

Introns: Non-coding regions of DNA within genes that initially were identified as non-essential, but some of which were later shown to have regulatory functions.

Phenotype: Observable physical characteristics of an organism, which depends on the genotype, are referred to as phenotype.

Proteomics: Study of the complete set of proteins produced by a cell.

Silent mutations: Changes in one of the letters of a codon, which does not alter the amino acid.

Transcription: The process of copying genetic information from the DNA template to RNA is called transcription.

Transcriptomics: Study of the complete set of all RNA transcripts produced by a cell.

Translation: The process of protein synthesis from mRNA using ribosomes within a cell.

Homologous recombination (HR): A type of genetic recombination in which nucleotide sequences are exchanged between two similar or identical molecules of double-stranded or single-stranded DNA in a cell.

Homology-directed repair (HDR): A mechanism in cells to repair double-strand DNA lesions by HR. HDR mechanism occurs in cells only when there is a homologous piece of DNA is present in the nucleus, mostly in G2 and S phase of the cell cycle.

Non-homologous end joining (NHEJ): NHEJ) is an alternate pathway present in cells to repair double-strand breaks (DSBs). During NHEJ, the DSBs are directly ligated without the need for a homologous template as opposed to HDR, which requires a homologous donor sequence.

Insertion-deletion (indels): Cellular repair of DSBs by NHEJ that is error-prone, that results in insertion or deletion of bases (aka indels) in the genome of an organism.

Suggested Reading

[1] Y-G. Kim, J Cha and S Chandrasegaran, Hybrid restriction enzymes: Zinc finger fusions to FokI cleavage domain, *Proc. Natl. Acad. Sci.*, USA, Vol.93, pp.1156–1160, 1996.

[2] J Smith, M Bibikova, F Whitby, A R Reddy, et al., Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger recognition

Address for Correspondence
Alok Kumar Singh
William R. Bishai Laboratory
Tuberculosis Research Center
Johns Hopkins School of
Medicine
1550 Orleans Street, CRB2,
Room 176
Baltimore, MD 21287, USA.
Email: aloksinghbharadwaj@
gmail.com
asingh51@jhmi.edu

Sivaprakash Ramalingam
CSIR-Institute of Genomics
and Integrative Biology
Room 330, Mathura Road,
Sukhdev Vihar
New Delhi 110 025, India.

Desirazu N Rao
Department of Biochemistry
Division of Biological Sciences
Indian Institute of Science
Bangalore 560 012, India.
Emial: dnrao@iisc.ac.in

Srinivasan Chandrasegaran
Department of Environmental
Health and Engineering
Johns Hopkins School of
Public Health
615 North Wolfe Street,
Baltimore, Maryland 21205,
USA.
Email: schandra@jhsph.edu

domain, *Nucleic Acids Res*., Vol.28, pp. 3361–3369, 2000.

[3] M Bibikova, D Carroll, D J Segal, J K Trautman, et al., Simulation of homologous recombination through targeted cleavage by a chimeric nuclease, *Mol. Cell Biol.*, Vol.21, pp.289–297, 2001.

[4] S Chandrasegaran and D Carroll, Origins of programmable nucleases for genome engineering, *J. Mol. Biol.*, Vol.428, pp.963–989, 2016.

[5] D A Wah, J A Hirsch, I Schildkraut, A K Aggarwal, Structure of Fok I has implications for DNA cleavage, *Proc. Natl. Acad. Sci.*, USA, Vol.95, pp.10564–10569, 1998.

[6] N P Pavletich, C O Pabo, Zinc finger-DNA recognition: Crystal structure of a zif268-DNA complex at 2.1 Å, *Science*, Vol.252, pp.809–817, 1991.

[7] M J Moscou and A J Bogdanov, A simple cypher governs DNA recognition by TAL effectors, *Science*, Vol.326, p.1501, 2009.

[8] J Boch, H Scholze, S Schomack, A Landgraf, et al., Breaking the code of DNA binding specificity of TAL-type III effectors, *Science*, Vol.326, pp.1509–1512, 2009.

[9] M Christian, T Cermark, E L Doyle, C Schmidt, et al., Targeting DNA double-strand breaks with TAL effector nucleases, *Genetics*, Vol.186, pp.757–761, 2010.

[10] J A Doudna and E Charpentier, The new frontier of genome engineering with CRISPR-Cas9, *Science*, Vol.346, pp.1077–1086, 2014.

[11] M Jinek, K Chylinski, I Fonfara, M Hauer, et al., A programmable dual RNA-guided DNA endonuclease in adaptive bacterial immunity, *Science*, Vol.337, pp.816–821, 2012.

[12] G Gasiunas, R Barrangou, P Horvath, V Siksnys, Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria, *Proc. Natl. Acad. Sci.*, USA, Vol.109, pp.e2579–2586, 2012.

[13] L Cong, F A Ran, D Cox, S Lin, et al., Multiplex genome engineering using CRISPR/Cas Systems, *Science*, Vol.339, pp.819–823, 2013.

[14] M Hoban, G Cost, M Mendel, Z Romero, et al., Correction of the sickle cell disease mutation in human hematopoietic stem/progenitor cells, *Blood*, Vol.125, pp.2597–2604, 2015.

[15] D Dever, R Bak, A Reinisch, J Camarena, et al., CRISPR-Cas9 *beta*-globin gene targeting in human haematopoietic stem cells, *Nature*, Vol.539, pp.384–389, 2016.