Welcome to the Brave New World: CRISPR Mediated Genome Editing-pathway to Designer Babies?

Diana Elizondo, Lourds Michelle Fernando, Everett Oliver1, Keely Clinton, Nicole Retland, Henri Paturault and Hemayet Ullah*

Department of Biology, Howard University, 415 College St., NW, Washington, DC 20059

Prologue
Quote from A Scientist about 25 years from now:
"Instead of creation of offspring by chance we can create offspring by choice."

Introduction
The world literature circle was abuzz in 1932 when Aldous Huxley published his seminal book “Brave New World”. He painted the efforts of a totalitarian state to lab manufacture “sub-human” people who would be capable of work but not of independent thought. Though the plot was set almost 500 years in the future, the author may not have envisioned that within a century of his writing scientists would embark on developing technologies that could potentially set the road for "designer babies" in the future. The recent availability of the simple, yet highly effective, CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) technology -- which is taking the scientific world by storm – provides the preliminary steps in the genome editing (making specific changes at targeted genomic sites) efforts of any biological organisms. The recent report of Chinese scientists’ use of the CRISPR system to rectify the mutation in a gene that causes beta-thalassemia disease in human embryos highlights the progression the scientific world is making in the realm of genome editing in germline (Liang et al. 2015). The scientific community has now widely accepted the fact that it is not a matter of if, but when, the Nobel Prize will be awarded for the development of this groundbreaking technology.

*Author for correspondence: <hullah@howard.edu>. 1Department of Biochemistry and Molecular Biology, Howard University, 420 W. St., NW, Washington, DC 200592.
Background on Genome editing tools

The ability to manipulate genes is important in elucidating their functions. The knowledge gained from these studies can be applied to treating diseases, such as alleviating certain metabolic defects, or to improve the quality of offspring, like in modifying plants to have higher crop yields.

Endonuclease-based targeted genome editing methods that emerged in the past decade are revolutionary molecular biology tools that have gained the attention of researchers around the world. Zinc Finger Nucleases (ZFNs), Transcription Activator-Like Endonucleases (TALENs) and CRISPR/Cas9 systems are programmable site-specific nucleases. Each of the nucleases act by inducing Double Strand Breaks (DSB) in DNA and result in the activation of error-prone Non-Homologous End Joining (NHEJ) and/or Homology Directed Repair (HDR) at the originally targeted genomic locus (Gai et al. 2013). Activation of the NHEJ or HDR process is then utilized to induce diverse genome manipulations in a sequence-specific manner, such as gene activation/inactivation, sequence deletion, element replacement and chromosomal rearrangement. Unlike previous gene therapy tools that add or insert an exogenous DNA copy into the target cell nucleus or genome, which may give rise to side effects such as insertional mutations and non-physical expression of proteins, programmable nucleases use a ‘cut-and-paste’ strategy to remove the defect and install the correct version (Xiao-Jie et al. 2015). This represents a preferable tool for gene therapy. The simplicity of the system is based on the introduction/injection of a Guide-RNA (gRNA) containing the target gene sequence (20 bp) and an endonuclease (Cas9) protein or mRNA into the cell. The injected molecules induce double strand DNA breaks in the targeted gene and the cell then activates NHEJ or HDR to repair the breaks and during the repair process induces the desired modification in the gene (see the schematic below). Compared to the protein based systems, the CRISPR/Cas9 system is more specific, convenient, efficient and is able to induce or rectify mutations in multiple places simultaneously in a wide range of organisms, including the human embryo. CRISPR/Cas9 has quickly become the most powerful genome engineering tool available in the field of biomedical as well as Plant Research.

Historical perspective of CRISPR

CRISPR/Cas system was discovered in bacteria as their adaptive immune response mechanism against foreign DNA such as viral DNA. CRISPRs are clustered genetic elements in the bacterial genome that contain parts of viral DNA acquired form the past viral infections. These snippets of DNA are called ‘spacers’ (Fig. 1). CRISPR was first described in E. Coli cells by Ishino and
Shinagawa (1987). Ishino discovered 14 repeating sequences which were regularly spaced but were random in sequence. The random arrays, along with conserved upstream gene sequences, were later found in a vast array of bacterial cells. By 2005 it was reported that the spacers often contained exogenous DNA, specifically, plasmid or phage DNA, resulting in the hypothesis that CRISPR played a role in bacterial immunity due to the fact that the invading phages are

![Diagram of CRISPR system](image.png)

**Fig. 1.** Schematic of the type II CRISPR-mediated DNA double-strand break in invading DNA. The type II CRISPR locus contains a cluster of four genes, Cas9, Cas1, Cas2, and Csn1, as well as two noncoding RNA elements, tracrRNA and a characteristic array of repetitive sequences (direct repeats) interspaced by short stretches of non-repetitive sequences (spacers, 30 bp each). Each spacer is typically derived from foreign genetic material (protopspacer) and drives the specificity of CRISPR-mediated nucleic acid cleavage. In the target nucleic acid, each protospacer is associated with a protospacer adjacent motif (PAM) whose recognition is exclusive to individual CRISPR systems. Type II CRISPR system carries out targeted DNA double-strand break (DSB) in sequential steps. Step 1, the pre-crRNA array and tracrRNA are transcribed from the CRISPR locus. Second, tracrRNA hybridizes to the direct repeats of pre-crRNA and associates with Cas9 as a duplex, which mediates the processing of the pre-crRNA into mature crRNAs containing individual truncated spacer sequences. Third, the mature crRNA/tracrRNA duplex directs Cas9 to the DNA target consisting of the protospacer and the requisite PAM via heteroduplex formation between the spacer region of the crRNA and the protospacer DNA. Finally, Cas9 mediates cleavage of target DNA upstream of PAM to create a DSB within the protospacer thereby inactivating any invading virus. (Adapted from www.biomart.cn).

recognized based on the similarity of their DNA sequence to the spacer sequences (Bolotin et al. 2005, Mojica et al. 2005). The hypothesis of a role in bacterial immunity was ultimately confirmed by evidence that there was a direct correlation between viral infection of bacteria and the number of spacers (Magadán et al. 2012). When a virus inserts its DNA into the bacteria, the transcribed ‘spacers’, which are about 60 nucleotides long RNA molecules complimentary to the piece of viral DNA, will bind to the invading viral DNA.
...and by using the endonuclease-like Cas9 (CRISPR associated protein 9) it can make double strand breaks in the targeted DNA thereby destroying the viral DNA. Scientists realized that by expressing DNA spacers specific to any gene sequences and the corresponding endonuclease like Cas9, a double stranded cut can then be induced in the target gene. The cells then try to repair the DSB by the error prone NHEJ or by HDR process. Both of the repair mechanisms can be utilized to introduce/rectify mutations in the double stranded cut DNA without the need for the spacers DNA to get integrated in the genome of the target organism. Researchers quickly realized how beneficial this natural defense mechanism could be if it was adapted for experimental purposes and many set out to determine the exact method by which CRISPR/Cas9 worked.

**Mechanism of CRISPR mediated genome editing**

The CRISPR defense mechanism as described above consists of three classes, all based on the CRISPR associated proteins (Cas proteins). The type II class is the most widely studied and what will be described in this review. By 2012 scientists became intimately aware of an adaptive defense system in certain bacteria (Magadán et al. 2012). At the time it was known that upon detection of exogenous DNA, bacterial cells responded by degrading the foreign/invading DNA from virus or plasmid. The resulting fragments (spacer) were then incorporated into the bacteria’s genome, essential for the recognition of repeat invasions. One or more existing conserved spacers are lost to prevent CRISPR to expand unchecked (Horvath et al. 2008, Tyson et al. 2008).

A repeat invasion by a virus or plasmid results in the transcription of the type II CRISPR locus, producing CRISPR associated proteins, Cas9 in particular, and two single-stranded RNA transcripts (Fig. 1). One of the RNA transcripts, is referred to as a precursor CRISPR RNA (pre-crRNA) and it contains the spacer sequence, therefore, it is complementary to the protospacer region (target region) of the exogenous DNA. The other transcript, tracer RNA (tracrRNA) binds to a complementary repeat sequence on the pre-crRNA. This newly formed double-stranded RNA prompts RNaseIII to cleave the pre-crRNA, producing a mature crRNA/tracrRNA complex. The crRNA/tracrRNA complex binds with the Cas9 nuclease to create a tracrRNA/crRNA/Cas9 complex.

A short sequence referred to as a protospacer-adjacent motif (PAM) located on the immediate 3’ end of the guideRNA (tracrRNA/crRNA together) recognition sequence also helps with target recognition. The PAM sequence varies by the species of the bacteria from which the Cas9 was derived. The most widely used Type II CRISPR system is derived from S. pyogenes and the PAM sequence is NGG located on the DNA strand complementary to the gRNA...
bound DNA strand. In addition to PAM, a short 12 to 15 nt sequence (seed sequence) upstream of the PAM sequence also helps with the recognition specificity. It is the job of Cas9 to scan the DNA for a PAM sequence and unwind the DNA when the PAM sequence is located. If there is complementation between the crRNA and the DNA, the complex will bind the DNA about 3 base pairs from the 5’ end of the PAM sequence and then Cas9 creates a double strand break (DSB) in the target DNA to silence the exogenous DNA. This nicked DNA can be repaired either by NHEJ or by HDR. NHEJ repair pathway results in insertions or deletions (InDels) where double strand breaks occurred which leads to a disruption in the open reading frame of the target locus. The HDR pathway needs a repair DNA template to fix the DSB and when a repair template is present, the result is integration of the sequence of DNA on the repair template into the DSB site, thereby replacing the genomic DNA in that region (Figure 2). Additional studies led scientists to realize that the programmed crRNA/tracrRNA complex could be replaced by an artificially programmed chimeric RNA, referred to as a guide RNA (gRNA) and that to accomplish genome editing, a cell could be simultaneously transfected/injected with the gRNA, the Cas9 nuclease, and in case of HDR - a donor DNA template.

Before attempting to explore Figs 2 and 3, which describe the detailed mechanism of the CRISPR process, the readers are advised to view the excellent animation on CRISPR developed by the McGovern Institute for Brain Research at MIT (https://www.youtube.com/watch?v=2pp17E4E-O8). Fig. 2 presents the generalized scheme of CRISPR-Cas9 mediated gene editing in a simplified way. After the gRNA/Cas9 complex is introduced in the cell, the complex is recruited to the target sequence by the base-pairing between the gRNA sequence and the complementary genomic DNA target sequence. For successful binding of Cas9, the genomic target sequence must also contain a Protospacer Adjacent Motif (PAM, blue region in Fig. 2A) sequence immediately following the genomic target sequence. The binding of the gRNA/Cas9 complex localizes the Cas9 to the genomic target sequence such that the wild-type Cas9 can cut both strands of DNA causing a Double Strand Break (DSB). Cas9 will cut 3 - 4 nucleotides upstream of the PAM sequence. Once cut, the DSB can be repaired through one of two general repair pathways: (1) the Non-Homologous End Joining (NHEJ) DNA repair pathway or (2) the Homology Directed Repair (HDR) pathway. The NHEJ repair pathway often results in inserts/deletions (InDels) in the DSB site that can lead to frameshifts and/or generation of premature stop codons, effectively disrupting the open reading frame (ORF) of the targeted gene. The HDR pathway requires the presence of a repair template, which is used to fix the DSB. HDR faithfully copies the sequence of the repair template to the cut target
sequence. Specific nucleotide changes can be introduced into the targeted genomic loci by the use of HDR with an appropriate modified repair template.

Fig. 2. Generalized scheme of CRISPR-cas9 mediated gene editing. After the introduction of gRNA containing complementary sequences to a targeted gene and Cas9 in the nucleus of target organism, the gRNA binds to the target sequence (A), the target DNA is cut by the Cas9 endonuclease (B). If no template DNA is available, NHEJ method is used by cells to repair the double stranded cut by incorporating/deleting random nucleotides in the cut site—resulting insertion/deletions in the target DNA (C1). If a donor DNA with a sequence that is intended to be incorporated in the cut site is used as a template, the cell will use the template DNA to do homologous recombination (as the donor DNA will have left and right flanking homologous DNA sequence) resulting in the replacement of the cut DNA with the donor DNA effecting the intended changes (C2).

In order to maximize the effect of gene disruption, target sequences should be chosen near the N-terminus of the coding region of the gene of interest. Typically, the target sequence is selected to introduce a DSB within the first or second exon of the gene. It is important not to design targets to introns (non-coding regions), as repair of the DSB in that region will not disrupt the target gene. When designing a repair template for genome editing by HDR, it is important that the repair template does NOT contain the target sequence followed by the PAM sequence or the template itself will also be cut by the Cas9. Changing the sequence of the PAM in the repair template should be sufficient to ensure it is not cut by Cas9. Finally, it is imperative to note that the changes introduced by use of the CRISPR system are permanent to the genomic DNA of
the organism (more can be found here: https://www.addgene.org/CRISPR/guide/).

![Diagram of CRISPR genome editing](image)

**Fig. 3**. Introduction of gRNA with sequences complementary to the 3'-CTCGATTATT-5' along with cas9 protein in the cell induces a double stranded cut at 3 - 4 bp upstream of the PAM sequence (TGG)- shown with the scissors. Cells try to use error-prone NHEJ to repair the double strand cut in the DNA. This repair mechanism does not use any template DNA to recombine but inserts (G/C in the figure) or deletes (AA/TT) random sequences in the cut to fill it up- effectively making the wild type gene sequence disrupted (left panel). In case, there is a need to rectify a mutation in a gene (C in number 4th position within the complimentary sequence to T in the corrected version), one will have to provide a donor DNA (single or double stranded) with the corrected sequence flanking left and right homologous sequences of the gene. Due to the homology of the left and right flanking sequences, the donor DNA will be used for homology recombination by the cell to fill up the double stranded cut in the DNA. The homology recombination will result in the exchange of DNA with the mutated sequence with the donor DNA containing the adjusted sequence- resulting in the correction of mutation (right panel).

Fig. 3 represents a hypothetical gene sequence that is disrupted by a deletion or insertion of random nucleotides (left panel) or uses a donor DNA to rectify a specific mutation (right panel). The general steps one would take to edit a gene are to first identify a 23 bp genomic site of the form 5'-N_{20}NGG near the intended target site (ideally +/- 50 bp). These bases may reside on the plus or on the minus strand. Using NCBI BLAST program, make sure your selected 23 bp genomic sequence is unique in the genome and that there are preferably none or very few sequences similar to the selected sequence at other locations in the genome. Putative target sites can also be identified by using an online tool such as ZiFit (http://zifit.partners.org/ZiFiT/ChoiceMenu.aspx) or CRISPR design (http://crispr.mit.edu). If you choose to scan for sites manually, you should still check your gRNA design for potential off-target effects by using ZiFit, CRISPR design.
tool, or another algorithm designed to predict off-target binding and mutagenesis. Scientists try to avoid gRNAs that have potential off-target sites with fewer than 3 mismatched bases (or fewer than 2 mismatched bases for a shorter, 17 bp guide). Once the 23 bp site has been identified, it is incorporated into a plasmid containing other conserved components needed for CRISPR to function, such as a U6 promoter, guide RNA scaffold, and terminator sequence. The whole sequence can be synthesized commercially by a company, like IDT, and cloned in a plasmid backbone. The resulting plasmid can be used to transcribe gRNA in vitro, and the gRNA and Cas9 protein can now be injected into the cell intended for editing. In the example of Fig. 3, once the gRNA and Cas9 are introduced, the Cas9 makes a double stranded cut about 3 bp upstream of the TGG PAM sequence (scissors cut in the figure). A cell that attempts to repair the double stranded cut with the NEJH pathway may result in a small genomic deletion (AA/TT in this example) or a new insertion (G/C). Either of these activities would disrupt the open reading frame—effectively disrupting the target gene. In the case of rectifying a mutation in the genome (for example, in Fig. 3, altering mutant (C/G) to wild type (T/A), a donor DNA (single or double stranded) with the wild type sequence flanking homologous sequences of the target gene is introduced. The cell can then attempt to repair the double stranded cut with the availability of the template DNA (donor DNA). Using homology based recombination, the template DNA with the wild type sequence replaces the cut site DNA—restoring the mutation to the wild type sequence.

Application of CRISPR

CRISPR application is currently leaning toward genome editing. Many diseases are genetic in nature, be it inheritable or acquired genetic mutations. CRISPR appears to be an effective technique that will allow scientists to adequately edit genes to cure diseases. Sickle cell anemia is a great example of a disease in which mutation of a single base mutation (T to A) could be edited by CRISPR and the disease cured. The case is similar for plant species, where scientists desire, for example, to knock-out a gene that will result in an increase in a particular nutritional content or in increased drought and/or pest resistance. Prior to the discovery of CRISPR, genome editing depended mainly on zinc finger nucleases and TAL effector nuclear (TALENs) techniques that are more laborious and lack efficacy. With the CRISPR method scientist now have the ability to use vectors containing, for example, a gRNA expression cassette with a suitable RNA polymerase III promoter (U6 or U3) to edit targeted gene loci. Thus far, there has been some success in this regard. Restriction enzyme site loss assays have been used to confirm successful CRISPR-mediated genome editing in model plants.
like *Arabidopsis thaliana* (Belhaj et al. 2015, Bortesi and Fischer 2015). Likewise, in an effort to confirm that gene editing was at least possible, cells from rice plants were transformed with vectors carrying CRISPR gateway vector targeting *CHLOROPHYLL A OXYGENASE 1 (CAO1)* gene (Miao et al. 2013). These examples at the very least confirm that a particular gene of interest was successfully targeted. More importantly, it has been shown in *Arabidopsis*, rice and tomato plants that the genetic changes induced by Cas9/gRNA were present in the germline and segregated normally in subsequent generations without further modifications (Bortesi and Fischer 2015). Ebina et al. (2013) reported that CRISPR-Cas9 can mutate long terminal repeat (LTR) sequence of HIV-1 in *vitro*, resulting in removal of the integrated proviral DNA from the part of the host cells and a significant drop in virus expression. Chronic hepatitis B is one the most common infectious diseases world-wide, which can lead to liver cirrhosis and cancer. Recently, a study by Seeger and Sohn (2015) showed that in HepG2 cells expressing hepatitis B virus (HBV), the introduction of CRISPR-Cas9 system resulted in both decreased hepatitis B core antigen expression which provides an impetus for further research on the possibility of CRISPR-Cas9-mediated hepatitis B prevention. In a mouse model of Duchenne muscular dystrophy (DMD), an inherited X-linked monogenic disease, CRISPR-Cas9-mediated gene editing in the germline gave rise to genetically mosaic offspring with 2-100% somatic cells carrying the corrected version of the culprit gene (Long et al. 2014). Studies by Wu et al. (2013, 2015) serves as another example where germline-manipulation with CRISPR-Cas9 system in mice were capable of correcting both the mutant gene and cataract phenotype in offspring initially caused by a one base pair deletion in exon 3 of *Crygc (crystallin gamma C)* gene. In human intestinal stem cells collected from patients with cystic fibrosis, the culprit defective gene *CFTR (cystic fibrosis transmembrane conductance regulator)* was rectified by homologous recombination during CRISPR-Cas9 genome editing while the pluripotency was retained as demonstrated by formations of organ-like expansions in cell culture (Schwank et al. 2013).

With regard to genome editing in plants, the CRISPR application does not have the negative stigma that is associated with genetically modified organisms (GMO). In fact, GMOs are banned in most European countries out of concern of unforeseen side effects. However, unlike GMOs, CRISPR relies on endogenous cellular mechanisms to silence or edit genes. Therefore, technically gene editing under the CRISPR method is not classified as genomic modification- as no new DNA is integrated in the host genome. The injected/transfected RNA/DNA and protein get diluted by continuous cell division and eventually falls below the level of detection. Most of the edits take place using cells own repair machineries;
hence the resulting edited genome is not identified as genetically modified. The result is that relative to GMOs less regulation and wider acceptance of CRISPR editing is expected. However, scientists are actively discussing the ethical aspects of this technology. Rudolf Jaenisch, MD, [a founding member of the Whitehead Institute for Biomedical Research and, President of the International Society for Stem Cell Research] expressed discontent at the rush toward the use CRISPR for gene editing (Jaenisch 2015). Dr. Jaenisch acknowledged the benefits of the science but questioned the circumstances that would justify its use in the absence of more scrutiny. Furthermore, Dr. Jaenisch has requested a moratorium to discuss the social and ethical issues. His remarks were made a few days after Chinese researchers published studies in connection with the use of human embryos to study the effectiveness of CRISPR in correcting thalassemia, an inheritable disease that results in anemia (Liang et al. 2015).

Future directions
If the CRISPR technique is to get beyond public scrutiny and possible regulatory demands, its off-target mutations must be corrected or at least minimized to an appreciable extent. Perfection of the off-target mutations will go a long way to quell such scrutiny. Currently, there are a couple of techniques that scientists are using to improve target selectivity. First, Cas9 has two nuclease domains, namely, Cas9 HNH and Cas9 RuvC (Shen et al. 2014). Researchers are currently looking into nullifying one of the nuclease sites followed by transformation of cells with two Cas9s; one with Cas9 HNH inactivated and the other with Cas9 RuvC. The mutated Cas9 proteins are referred to as Nickase, as they form single-stranded nicks in the DNA as opposed to double-strand breaks. The technique requires two individual gRNAs. The logic behind the Nickase protocol is that single-strand nicks are easily repaired without the creation of the mutations associated with double-strand break repairs (Fig. 3). Similarly, in 2014 Doudna - co-founder of the type II CRISPR technique indicated that conformational changes occur in Cas9 upon binding to the gRNA, with additional conformational changes upon binding to double-stranded DNA (Doudna and Charpentier 2014). Doudna speculated that modifications in the gDNA and/or Cas9 could affect conformational changes enough to render Cas9 inactive upon off-target binding, thereby reducing off-target mutations. Advances have also been made by using completely inactivated Cas9s in order to perform an array of techniques from physically blocking gene expression, by acting as a signaling system, etc. The inactivated Cas9 is referred to as dead Cas9 (dCas9) and are used in this regard (Agne et al. 2014). Dead Cas9 have been inactivated such that they no longer possess nuclease activity. When a dCas9 is paired with a gRNA,
the complex silences gene expression by simply base pairing with the target gene, without cutting the DNA. Relative to RNAi, gene silencing with dCas9 is more pronounced and takes place over a longer time period. DCas9 can also be used as a signaling probe to locate and map specific genes. Of course, the future direction of CRISPR has been enhanced with the creation of gRNA libraries and computer programs to help with gRNA programming.

Acknowledgement

This work culminates from a biology class exercise in the Department of Biology, Howard University, Washington, DC, USA. The authors greatly acknowledge the service of Dr. Anna Allen for the critical readings of the manuscript. In addition, they recognize the excellent editing service of Aniqa Tasnim as well.

References


Boilotin A1, Quinquis B, Sorokin A and Ehrlich SD (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology 151: 2551-61.


