16 - Genome Editing

Genome editing allows scientists to introduce targeted changes to the DNA of an isolated cell or an entire organism. For example, scientists can insert a DNA sequence, delete a DNA sequence, or modify the DNA sequence of any gene via genome editing. The goal of genome editing is to change the phenotype of a cell in a way controlled by the researcher.

The applications of genome editing are staggering. Genome editing can be used in research to better understand the role of a gene and its protein product in cellular structure and function. Genome editing can also be used as a treatment for genetic diseases by replacing a mutant gene that causes the disease with the normal (wild-type) version of the gene. Finally, genome editing can be used to enhance the yield of crops or give desirable traits to livestock.

Key Questions

- What is meant by genome editing?
- What are some of the applications of genome editing?

Genome editing (overview)

Genome editing works by recognizing a specific **target DNA sequence** in the genome. After recognition of the target DNA sequence, an **endonuclease** cuts both DNA strands. The cell then tries to fix the double-stranded DNA break by rejoining the two ends of the severed DNA molecule; however, the repair mechanisms involved are error-prone, introducing extra nucleotides or deleting nucleotides at the cut site. The insertion or deletion of a single base or two bases within the coding region of a gene changes every codon downstream of this insertion/deletion (**indel**) site. This type of mutation, referred to as a **frameshift mutation**, produces a defective protein product (see Part 7).

Key Questions

- How does genome editing work?
- What is meant by an indel mutation and a frameshift mutation?

Genome editing systems

There are three major genetic technologies that can be used to edit DNA sequences within isolated cells or entire organisms:

- **Zinc-finger nucleases (ZFNs).** ZFNs are enzymes engineered in the lab to contain two parts: a **zinc-finger motif** and an endonuclease. The zinc-finger motif allows the ZFN to bind to the target DNA sequence (see Part 14). One ZFN attaches to one DNA strand at the target site; a second ZFN binds to the other DNA strand about ten base pairs away. The endonucleases come together and cut both DNA strands between the ZFN binding sites. Because the ZFN binds and then cuts a specific target DNA sequence, ZFNs only create a single genome edit at a time.
- **Transcription activator-like effector nucleases (TALENs).** Like the ZFNs, TALENs are enzymes designed by researchers to include both a DNA-binding region and an endonuclease region. The TALEN DNA-binding protein domain can be engineered to bind to any target DNA sequence. Once bound to the DNA, the TALEN endonuclease domain cuts both strands of the target DNA sequence, allowing the creation of a single genome edit at a time.
- **CRISPR-Cas9.** The CRISPR-Cas9 system is the newest, most powerful, and versatile genome editing technique. CRISPR-Cas9 can be used to create a single genome edit or multiple genome edits simultaneously.

ZFNs and TALENs have many drawbacks, including the high cost and time involved in engineering the DNA binding domains within the nucleases and the inefficient cutting of the target DNA sequence. Although the ZFNs and TALENs have been used to successfully edit genes, the science world has embraced CRISPR-Cas9 due to its lower cost, higher efficiency, and potential to create multiple genome edits simultaneously. Because of its widespread current use and promising future, CRISPR-Cas9 will serve as the subject for the remainder of this chapter.

Key Questions

- Describe the three major genome editing technologies.
- Why do scientists prefer CRISPR-Cas9?

The CRISPR-Cas9 genome editing system

CRISPR is an acronym for the **clustered regularly interspaced short palindromic repeats (CRISPR)** system. The CRISPR-Cas9 system has two molecular components (see **figure 16.1**):

- A **single guide RNA (sgRNA)** The sgRNA consists of a single-stranded RNA molecule called **crRNA** that forms hydrogen bonds with a specific target DNA sequence. The crRNA is covalently linked to a stem-loop RNA sequence called **tracrRNA.** The tracrRNA binds to and activates the Cas9 endonuclease to cut the double-stranded DNA at the target site.
- A **CRISPR-associated endonuclease protein (Cas)**. The Cas protein is a non-specific endonuclease that cuts double-stranded DNA when activated by the tracrRNA. The genome editing system described below uses the **Cas9** enzyme isolated from the bacterium *Streptococcus pyogenes*.

Additionally, the DNA sequence targeted by the sgRNA needs to contain a **protospacer adjacent motif (PAM)** sequence, as Cas9 binds to the PAM sequence to position itself while it cuts both strands of the DNA. The PAM sequence is a DNA consensus sequence consisting of 5'-NGG-3', where N is any of the four DNA bases (A, T, C, or G). The PAM sequence is in the **nontarget DNA strand**; the nontarget DNA strand does not form hydrogen bonds with the crRNA component within the sgRNA. The PAM in the nontarget DNA strand is located 3-4 nucleotides in the 3’ direction (downstream) from the site that will be cut by Cas9.

The CRISPR-Cas9 system creates genome edits as follows:

1. Cas9 binds to the PAM sequence in the nontarget DNA strand.
2. The target and nontarget DNA strands are separated from each other. The Cas9 enzyme is the helicase that separates the two DNA strands.
3. The crRNA attempts to form hydrogen bonds with the target DNA strand. If the crRNA forms proper hydrogen bonds with the target DNA strand, then genome editing continues. If hydrogen bonds fail to form, the Cas9 enzyme is released and binds to another PAM sequence in the genome.
4. The binding of the crRNA to the target DNA strand activates tracrRNA, which in turn, activates Cas9.
5. Cas9 enzyme cuts both DNA strands 3-4 nucleotides in the 5' direction (along the nontarget strand) from the PAM site.
6. Once both strands of the DNA have been cut by Cas9, the cell's DNA repair systems attempt to fix the break in the dsDNA and, in doing so, add a few bases, delete a few bases, or insert a completely new piece of DNA.

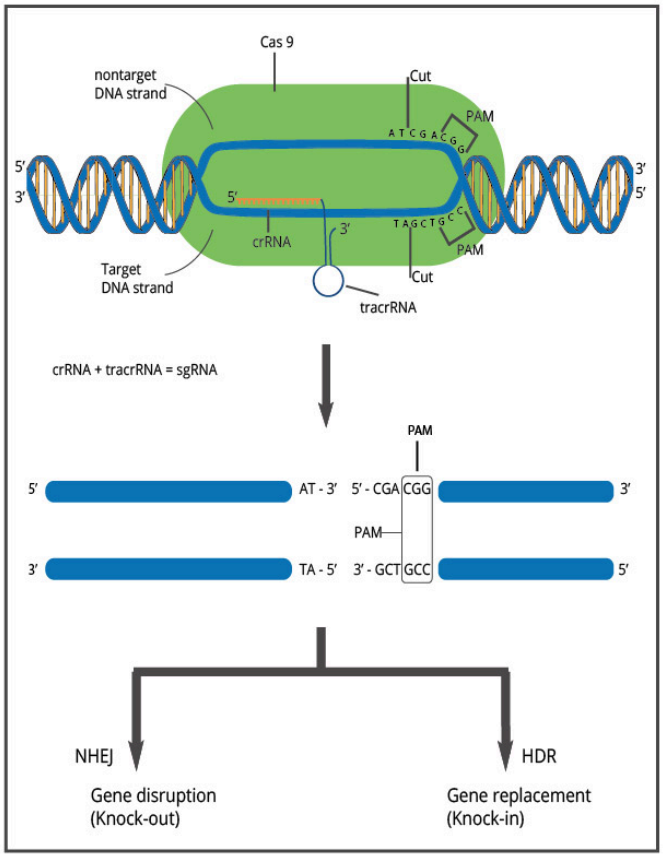

Figure 16.1 ***The CRISPR-Cas9 System*** *--- Image created by SL*

Key Questions

- What is meant by the target and nontarget DNA strands?
- What are the names of the two components within a sgRNA molecule?
- Describe how crRNA, tracrRNA, Cas9, and the PAM contribute to the CRISPR-Cas9 genome editing system.

What is the natural function of CRISPR-Cas9?

The CRISPR-Cas9 system is thought to be analogous to an immune system, protecting bacteria against invading bacteriophages (viruses that infect bacteria). During an infection, the bacteriophage genome is injected into the cytoplasm of the bacterial cell. The bacteriophage DNA is cut by nucleases, and a portion of the bacteriophage genome

is stored in the **CRISPR gene locus**. Overall, the CRISPR gene locus in bacteria consists of clusters of repetitive DNA sequences (short palindromic repeats that are 30-40 base pairs in length) separated by bacteriophage DNA sequences called **spacers**. In essence, the spacer sequences within the CRISPR locus are a library of previous bacteriophage infections (see **Figure 16.2**).

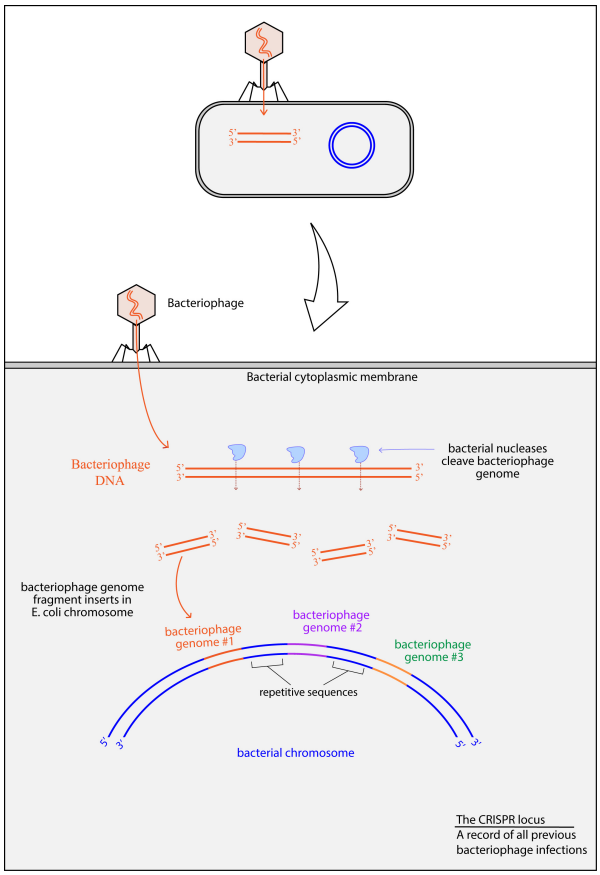

Figure 16.2 - The CRISPR locus is a library of previous bacteriophage infections. *Fragments of bacteriophage genomes are stored as spacers in the bacterial chromosome. Image created by Alex Baff.*

Upon reinfection with the same bacteriophage, the CRISPR gene locus is transcribed to produce two types of RNA molecules (see **Figure 16.3**). The spacer DNA sequence (i.e., the bacteriophage genome) is transcribed to produce the single-stranded CRISPR RNA **(crRNA)** to form hydrogen bonds with the DNA of the infecting bacteriophage. Another gene in the CRISPR locus is transcribed to make the **transactivating crRNA (tracrRNA)**. The crRNA and the tracrRNA from hydrogen bonds with each other and then bind to the Cas9 endonuclease. Note that the tracrRNA contains the stem-loop that activates Cas9. The crRNA:tracrRNA:Cas9 complex then binds to a PAM sequence in the DNA of the invading bacteriophage. The two DNA strands within the bacteriophage DNA are separated and the crRNA forms hydrogen bonds with the **target DNA strand**, while the **nontarget DNA strand** is moved out of the way. Finally, the Cas9 protein makes **double-stranded breaks** (DSB) in the DNA of the bacteriophage, thereby destroying the bacteriophage genome and inhibiting the bacteriophage infection.

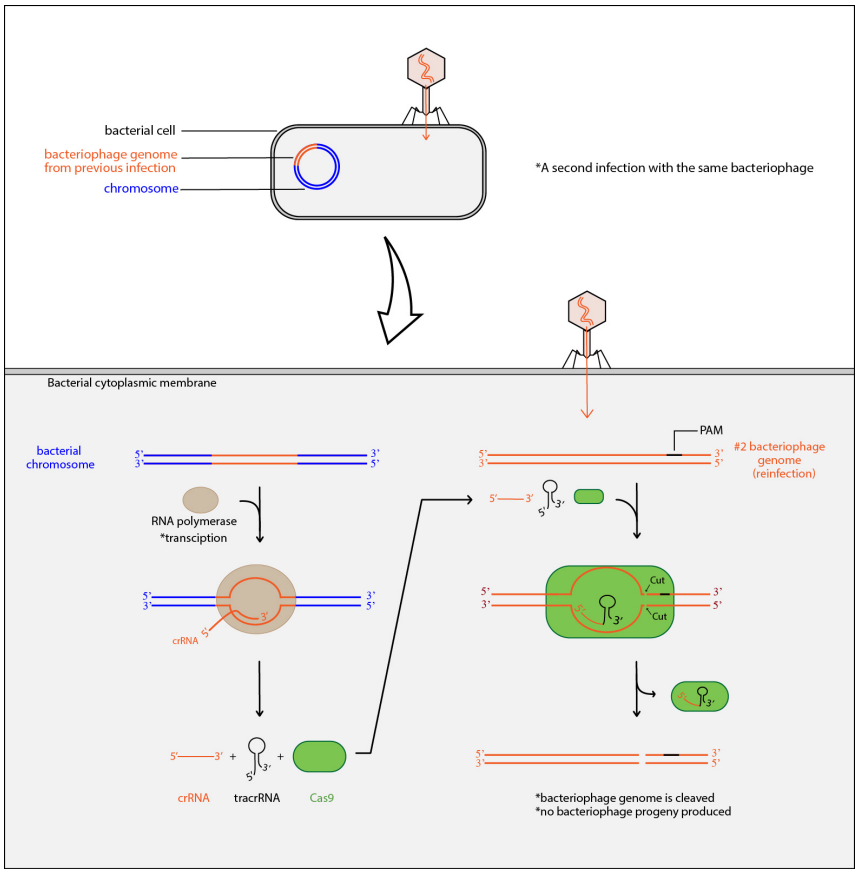

Figure 16.3 - The CRISPR-Cas9 destroys the bacteriophage genome upon reinfection. *Image created by Alex Baff.*

Key Question

- How is the CRISPR-Cas9 system beneficial to a bacterial cell?
- How are spacers generated?
- Describe how the CRISPR-Cas9 system destroys the DNA of an invading bacteriophage.

Applications of CRISPR-Cas9

Genome editing via CRISPR-Cas9 involves designing a 20 nucleotide-long crRNA sequence that forms hydrogen bonds with a target DNA sequence of interest. This crRNA is covalently linked to the tracrRNA that forms the RNA stem-loop to activate Cas9. The crRNA linked to the tracrRNA is the **sgRNA** component of the CRISPR-Cas9 system. Both the sgRNA and Cas9 DNA sequences are ligated into separate cloning sites within a plasmid vector, and the plasmid vector is introduced into a cell of interest, including a eukaryotic cell. Transcription of the cloned genes leads to the production of both the sgRNA and Cas9 molecules.

When the sgRNA binds to a target DNA sequence, Cas9 produces a **double-stranded break (DSB)** in the DNA. When the cell attempts to repair these DSBs, the cell can undergo the **non-homologous end joining (NHEJ)** DNA repair pathway. NHEJ is not perfect, and insertion or deletion of a few nucleotides occurs (these mutations are called **indels**) as the DSB is repaired. Recall that indels cause a frameshift during translation that ultimately prevents the eukaryotic gene

from making a functional protein product. Therefore, CRISPR-Cas9 genome editing followed by NHEJ allows the researcher to produce a gene **knock-out** cell line or organism. The knock-out fails to produce a functional protein product.

Double strand breaks in the DNA can also lead to another type of DNA repair known as **homology directed repair (HDR)**. In this case, DNA repair allows the insertion of a **donor sequence** at the location of the DSB, instead of repairing the break by inserting or deleting a few nucleotides. The donor DNA can be engineered to contain a mutant form of a gene. This approach allows the researcher to insert a mutant gene in the place of a wild-type gene to study the effects of the mutation on the cell. Alternatively, the donor DNA sequence can contain a wild-type version of a gene that replaces the mutant form of the gene within the cell. The replacement of a gene with a different allele of the same gene produces a **knock-in** cell.

CRISPR-Cas9 is a convenient genome editing system to use because if a scientist wishes to study a different gene, the scientist designs a new 20 nucleotide-long crRNA that forms hydrogen bonds with the new target gene, all of the other components (i.e., tracrRNA, Cas9) of the CRISPR-Cas9 system remain the same. Moreover, the use of multiple unique crRNA sequences allows the alteration of several genes in the genome simultaneously.

Key Questions

- How is gene cloning used in genome editing?
- Describe how NHEJ can be used to create a knock-out cell.
- Describe how HDR can be used to create a knock-in cell.

Challenges associated with CRISPR-Cas9

Before we explore the ethics of genome editing, let us investigate some of the challenges of using the CRISPR-Cas9 system. In a typical experiment, the researcher will introduce the CRISPR-Cas9 vector into a population of eukaryotic cells. Because the process of genome editing is inefficient, the experiment will result in three groups of cells in the population: those in which no editing occurred, those in which one of the two alleles of a gene is edited, and those with both alleles edited. If the knock-out approach is used, the researcher will want to study cells that have no functional copies of the gene; therefore, the researcher will need to identify those cells with both alleles edited. Determining the DNA sequence of the target gene in individual cells is one of the easiest ways to confirm that the desired changes have taken place.

The crRNA is designed to target a specific gene in the genome; however, sometimes a 20 nucleotide-long crRNA can bind to more than one DNA sequence in the genome simultaneously. This raises the possibility that the CRISPR-Cas9 system will cut the DNA at undesired locations within the genome, producing **off-target effects**. Because the locations of these off-target cut sites are difficult to predict, treating cells with CRISPR-Cas9 can have unintended consequences on the cell or organism.

Key Questions

- What are two challenges associated with CRISPR-Cas9 genome editing?

The ethics of genome editing

Many scientists are interested in using CRISPR-Cas9 to treat human genetic diseases, especially diseases for which there is currently no treatment. There are two main ways that human genome editing can be used to treat disease: inactivation of a mutant gene to remove its effects on the cell (using the NHEJ knock-out approach) or insertion of a functional allele to replace a mutant one (using the HDR knock-in approach).

With the promise that CRISPR-Cas9 brings, there is also uncertainty about the ethics of this technique, particularly when applied to humans. Most researchers agree that if we have the tools necessary to treat a genetic disease, we should use those tools to improve the lives of patients. However, considerable disagreement exists as to whether the CRISPR-Cas9 technique should be used to modify the germ-line cells that produce gametes or embryos. Current laws in the United States prohibit the use of human genome editing in gamete-producing cells. Research on human embryos is permitted if the treated human embryos are destroyed before day 14 of development and are not implanted into the womb.

An important issue to consider with genome editing is that of informed consent. An adult can give consent for genome editing that can potentially treat their genetic disease, but when that treatment extends to future generations, there is no way to obtain consent (i.e., the developing fetus cannot give consent). Public opinion remains divided as to who has the right to make the decision for the fetus; is it the person who develops from the embryo, parents, or the government?

In November 2018, the press announced that a researcher in China used CRISPR-Cas9 to successfully edit the *CCR5* gene in human twins (one received the edit while the other did not). Knocking out this gene is expected to prevent the treated child from contracting a human immunodeficiency virus (HIV) infection. To say the scientific world was upset about this announcement is an understatement. This was the first time that a human baby was born after genome editing was performed. The reason why this announcement was not received with congratulations was, in part, due to the lack of informed consent and the failure to make sure that no off-target effects took place before implanting the embryos into the birth mother. In fact, it is uncertain if the parents were informed as to the genome editing experiment, or if they were coerced into giving their consent.

There may never be an international agreement concerning genome editing that can be enforced by all nations. Even when there is agreement as to what is ethical and what is not, there will always be individuals or nations who will carry out research that is contrary to the moral beliefs of others. Important questions to consider include how do we establish laws concerning the ethical practice of scientific research, and how do we penalize those who knowingly disobey those laws?

In 1956, mathematician and biologist Jacob Bronowski wrote that, as scientists, “We ought to act in such a way that what is true can be verified to be so,” an expression of his belief that it is our right and our duty to explore the unknown and seek truth. The point is that maybe having large international committees decide what should be practiced and what should be prohibited is not the real question, but rather how can society ensure that research done is based on the priniciple of seeking truth to better the lives of humankind?

Key Questions

- Should genome editing be done on gamete-producing cells, embryonic cells, or somatic cells? Why or why not?