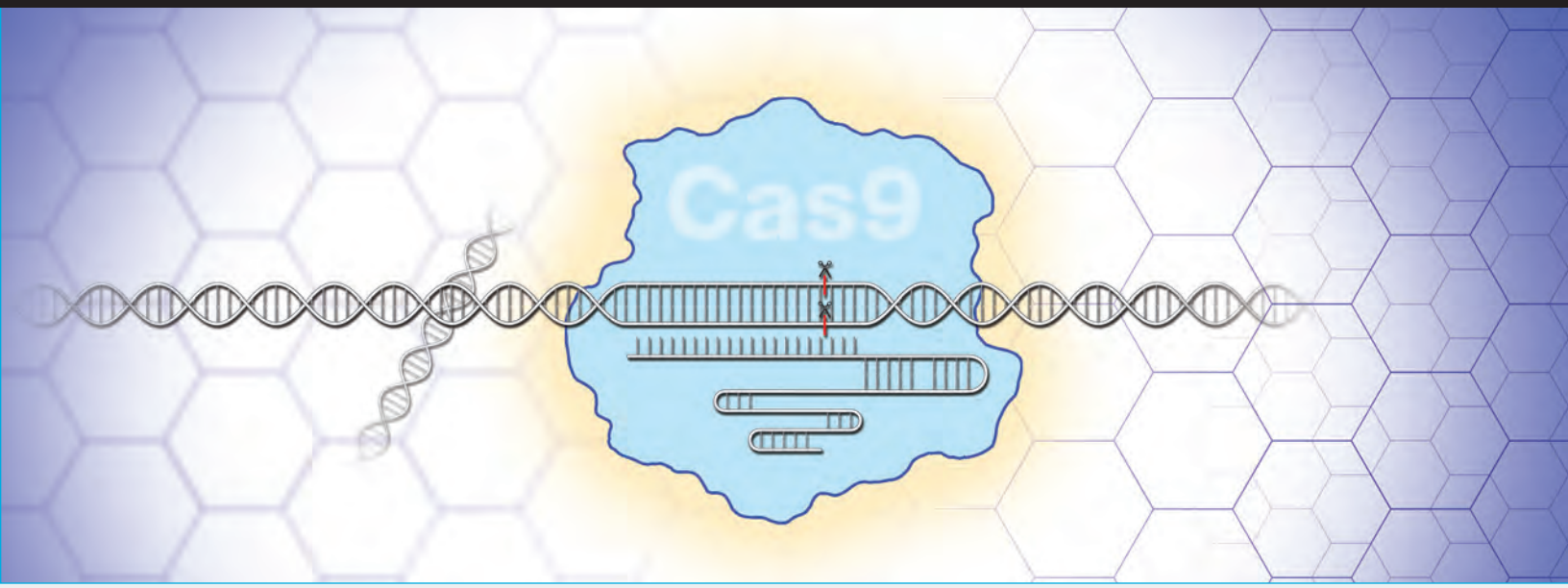
BIO-RAD Explorer



Out of the Blue CRISPR Kit

Catalog #12012608EDU

Student Guide

For technical support, call your local Bio-Rad office or, in the U.S., call **1-800-4BIORAD** (1-800-424-6723) option 2.

This product is not for self-administration.

BIO-RAD

This page is intentionally left blank

Activity 1

Introduction to CRISPR-Cas9 Gene Editing Technology

What is CRISPR-Cas9 gene editing?

In the decades since the discovery of restriction enzymes, researchers have discovered many new molecular tools and techniques that have greatly expanded our genetic engineering capabilities. One of the most exciting recent developments is the CRISPR-Cas9 system (CRISPR). CRISPR derives its name from the system found in nature that allows microbes to defend themselves against viral attack. "Clustered regularly interspaced palindromic repeats" (CRISPR) are sequences in the genomes of some prokaryotes that act as a genomic record of previous viral attack. Along with CRISPR-associated (Cas) proteins, bacteria use the sequences to recognize and disarm future invading viruses as shown in Figure 1. Scientists have adapted this system for genetic engineering purposes.

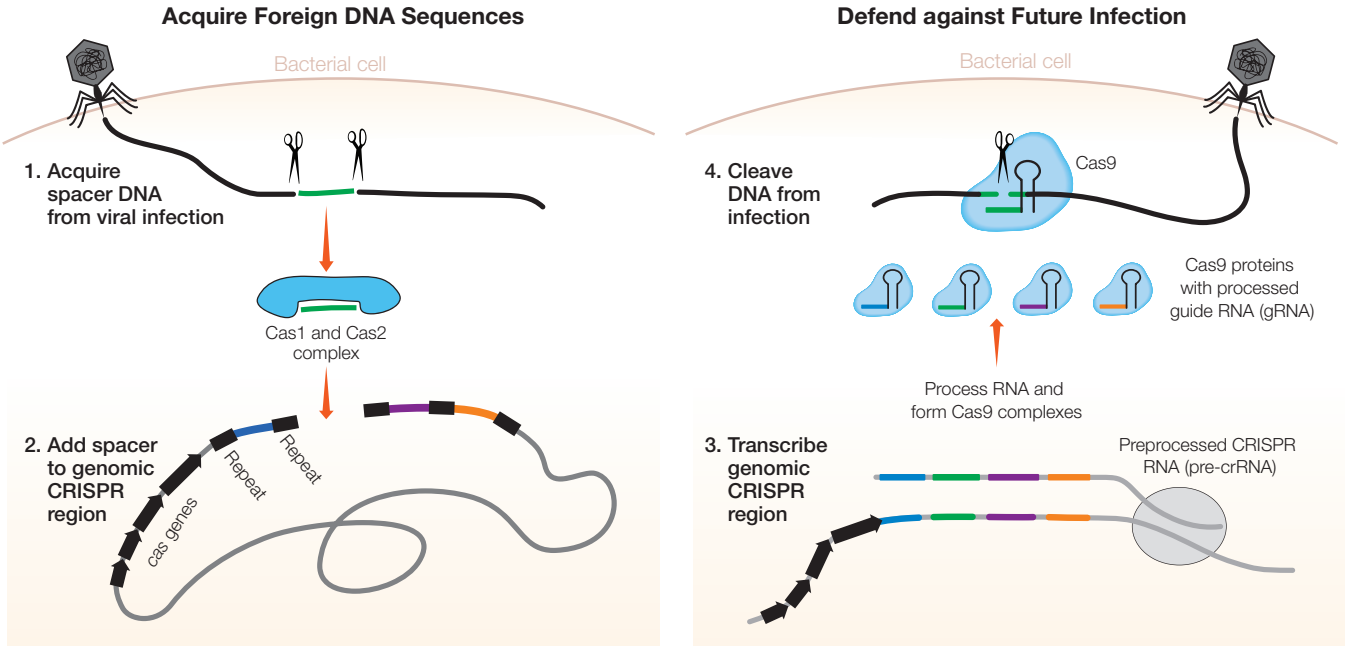


Fig 1. The CRISPR-Cas9 microbial defense system. 1, the Cas1-Cas2 enzymes of the microbe recognize foreign DNA and cut out a segment; 2, the Cas1-Cas2 enzymes insert the DNA segment into the CRISPR region of its own genome as a spacer; 3, a spacer sequence is transcribed and then linked to a Cas9 protein; 4, upon reinfection by the same invader, the CRISPR-Cas9 complex can recognize the foreign DNA sequence and cut it to prevent complete infection.

CRISPR-Cas9 is not the first programmable gene-editing tool, nor is it necessarily the most precise. Other gene-editing tools, like transcription activator-like effector nucleases (TALENs) or zinc-finger nucleases, are also programmable and precise, but they are very expensive and laborious to use. What makes CRISPR-Cas9 so powerful is the combination of its precision and simplicity.

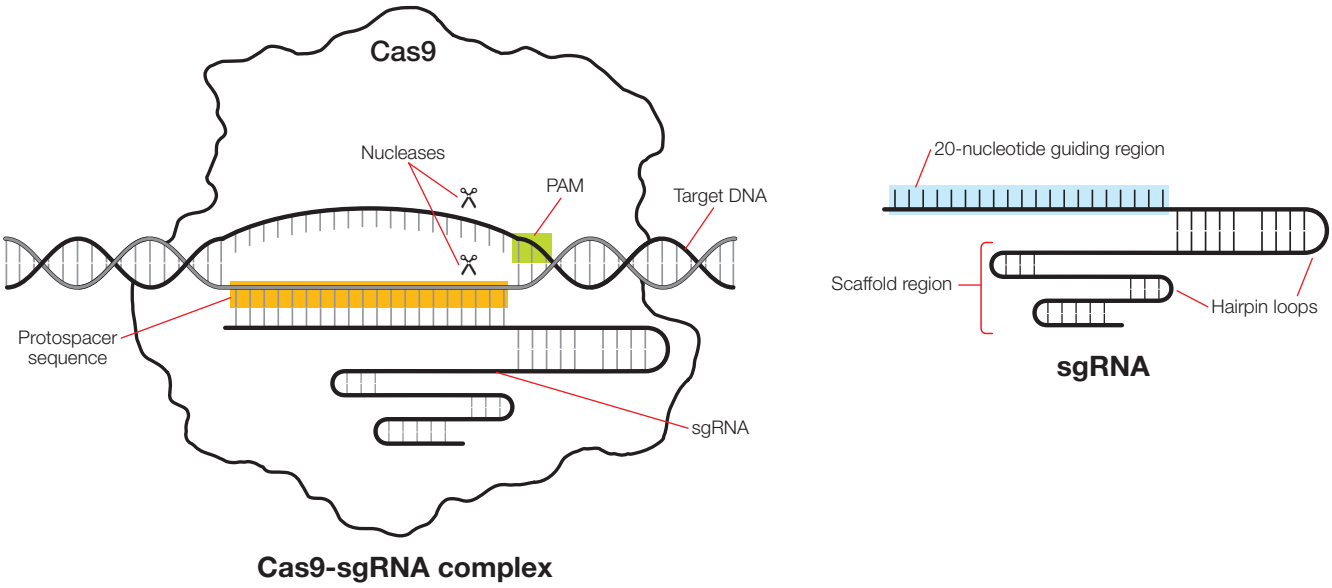


Fig. 2. Anatomy of Cas9 and sgRNA.

The CRISPR-Cas9 system consists of the following components as shown in Figure 2:

- **Cas9 enzyme (Cas9)** — a bacterial endonuclease that forms a double-strand break (cuts) DNA at a specific site within a larger recognition sequence, or target site. The Cas9 recognition sequence includes a 20-nucleotide sequence called the protospacer that is determined by a guide RNA bound to the enzyme
- **Single guide RNA (sgRNA)** — an engineered form of guide RNA that forms a complex with Cas9. The sgRNA is an approximately 100 nucleotide–long fusion of two regions that occur as separate RNAs in nature:
  - Guiding region — part of the CRISPR RNA or crRNA in nature, a typically 20-nucleotide region that is complementary to the target DNA sequence and that defines where Cas9 cuts. Scientists can easily customize this sequence for their own targets
  - Scaffold region — called the transactivating CRISPR RNA or tracrRNA in nature, a region that forms a multi-hairpin loop structure (scaffold) that binds tightly in a crevice of the Cas9 protein. The sequence of this region is typically the same for all sgRNAs
- **Protospacer adjacent motif (PAM)** — a sequence motif immediately downstream of the protospacer sequence in the Cas9 recognition sequence that is required for Cas9 function. Cas9 recognizes the PAM sequence 5'-NGG where N can be any nucleotide (A, T, C, or G). When Cas9 binds the PAM, it separates the DNA strands of the adjacent sequence to allow binding of the sgRNA. If the sgRNA is complementary to that sequence, Cas9 cuts the DNA

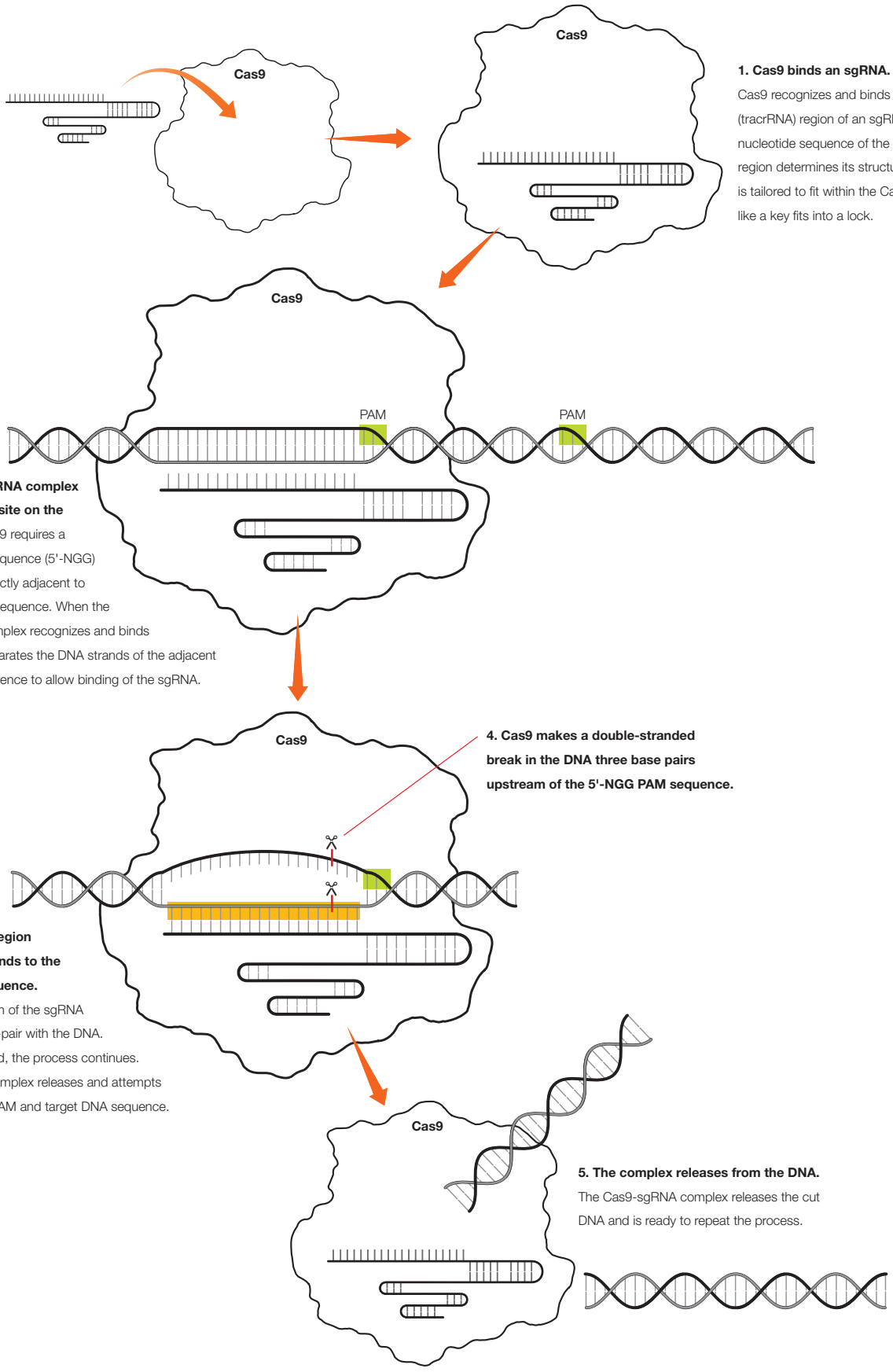


Fig. 3. The steps of Cas9 DNA recognition and cleavage.

Part 1. Simulate the Molecular Mechanism of Cas9 DNA Cleavage

Use the paper model to walk through the steps of CRISPR-Cas9 DNA cleavage using a sequence from the bacterial gene *lacZ* which encodes β-galactosidase. The *lacZ* gene is part of the lac operon, a collection of genes that allows bacteria to use lactose, a milk sugar, as a food source. The DNA and sgRNA sequences in the paper model match those used in Activity 2, *lacZ* CRISPR Gene Editing Laboratory.

1. ***Cut out the sgRNAs and DNA strips. You may leave the Cas9 protein on its page.***
2. ***Use the steps in Figure 3 as a guide to model the CRISPR-Cas9 mechanism:***
   a. Cas9 binds an sgRNA: Place sgRNA 1 onto the Cas9 illustration and align it with the dotted lines.
   b. The Cas9-sgRNA complex binds to a PAM site. Place DNA strip 1 on the stripe across the Cas9 model. Slide the DNA strip until the PAM box on the Cas9 protein matches a PAM (5'-NGG) sequence on the DNA.
   c. The guiding region of the sgRNA binds to the target DNA sequence. Check whether the DNA sequence is complementary to the sgRNA sequence (U pairs with A, C pairs with G). If they are complementary, continue the process. Otherwise, repeat steps 2.b and 2.c with a new PAM site.
   d. Cas9 makes a double-stranded break in the DNA: The scissors icons indicate where Cas9 cuts the DNA strands. Use a pencil to draw a vertical line across both strands at this position.
3. ***Verify that you have chosen the correct cut site and then use a pair of scissors to cut DNA strip 1 at that site. Keep the pieces of DNA strip 1 for use in Part 4.***

Focus questions

A. How many nucleotides long is the guiding region of the sgRNA?

B. Does the sgRNA bind to the PAM?

C. Where does Cas9 cut the target DNA relative to the protospacer sequence?

Part 2. Design the Guiding Region of an sgRNA

CRISPR technology is powerful in part because the target DNA sequence is controlled by a customizable sgRNA. In this activity, you will customize the guiding region of the sgRNA to cut a target site on the *lacZ* gene. DNA strip 2 represents a DNA sequence from the *lacZ* gene where you wish to make an edit.

1. ***Use sgRNA 2, DNA strip 2, and the steps you followed in Part 1 to determine the sgRNA guiding region sequence required to direct Cas9 to cut DNA strip 2 at the red dashed line.***
2. ***Write the nucleotide letters (A, U, C, G) of this sequence into the spaces on sgRNA 2.***
3. ***Use the steps of Cas9 DNA cleavage to confirm that the sequence you wrote on sgRNA 2 is correct.***
4. ***Record your final sequence below or in a notebook.***

sgRNA guiding region sequence

Focus questions

A. Describe in complete sentences how the requirement of a PAM sequence affects the flexibility of CRISPR-Cas9 gene editing.

B. Describe in complete sentences how you would identify a target DNA cleavage site for CRISPR-Cas9 and design an sgRNA.

Part 3. Compare the Specificity of DNA-Cutting Tools

The flexibility and specificity of CRISPR-Cas9 technology offer a large step forward for gene editing. The first DNA "scissors" were restriction enzymes, which cut DNA at predefined sequences, typically 4–8 base pairs long. For example, EcoRI, a restriction enzyme found in *E. coli*, will cut double-stranded DNA at every GAATTC sequence. If EcoRI were added to a sample that contained the entire human genome, it could cut at every GAATTC sequence.

We can calculate the probability that a particular nucleotide sequence, such as GAATTC, will occur within a larger sequence. Table 1 below shows the calculated probabilities of finding sequences of particular lengths. These calculations are based on the assumption that DNA sequences are entirely random and that every nucleotide position has an equal probability of being A, T, C, or G. Use the table to answer the following questions.

Table 1. Calculated probabilities of finding a specific sequence.

Sequence	Sequence Length	Probability Calculation	Predicted Occurrence in a Sequence the Length of the Human Genome (3,234,830,000 bases)
A	1	$¼ = (1/4)^1 = 0.25$	808,707,500
AC	2	$¼*¼ = (1/4)^2 = 0.0625$	202,176,875
GAATTC (EcoRI)	6	$¼*¼*¼*¼*¼*¼ = (1/4)^6 = 2.44 \times 10^{-4}$	789,753
NNNN…	n	$¼*¼*¼*...= (1/4)^n$	$(1/4)^n * 3,234,830,000$

Focus questions

A. What is the probability that any base in a sequence is an adenine, A? How many times do you expect to find adenine in the human genome?

B. What is the probability of finding a particular two-base sequence? How many times do you expect to find that sequence in the human genome?

C. How many times would you expect to find an EcoRI cut site in a fragment of DNA 1,000,000 base pairs long?

D. How many times would you expect to find a specific 20 base pair sequence in the human genome?

E. Write out a complete equation to calculate the predicted occurrence of a sequence of n length within a DNA fragment of X length.

F. Using mathematical evidence, explain why CRISPR-Cas9 gene-cutting technology, which uses a target sequence of 20 base pairs, is more specific than classic restriction enzymes.

G. Write three different ideas you have about why CRISPR-Cas9 technology could be more useful for gene therapy and/or research than other gene-cutting tools.

H. In actuality, the DNA sequence of the human genome is NOT random. Some sequences, including some very large sequences, are repeated many times throughout the human genome. Write two ideas you have for how this fact complicates the use of CRISPR gene-editing technology in humans.

Part 4. Design Donor Template DNA for DNA Repair

CRISPR-Cas9 can find a specific sequence in a genome billions of base pairs long and then cut at a precise location within that sequence. How do scientists and researchers use the specificity of CRISPR-Cas9 to direct targeted gene editing?

When chromosomal DNA in a bacterial cell is cut, the cell will die unless it's able to repair the cut. Bacteria have evolved processes to repair double-strand DNA breaks that would otherwise lead to cell death. DNA repair can happen in two ways, as shown in Figure 4:

- **Nonhomologous end joining (NHEJ)** — enzymes reconnect the ends of the double-stranded break back together. This process may randomly insert or delete one or more bases and can cause mutations that can disrupt gene function or expression
- **Homology directed repair (HDR)** — enzymes patch the break using donor template DNA. Researchers design the donor template DNA, which may include a desired sequence flanked on both sides by "homology arms" that match the sequence upstream and downstream of the cut. A complementary DNA strand is created during repair

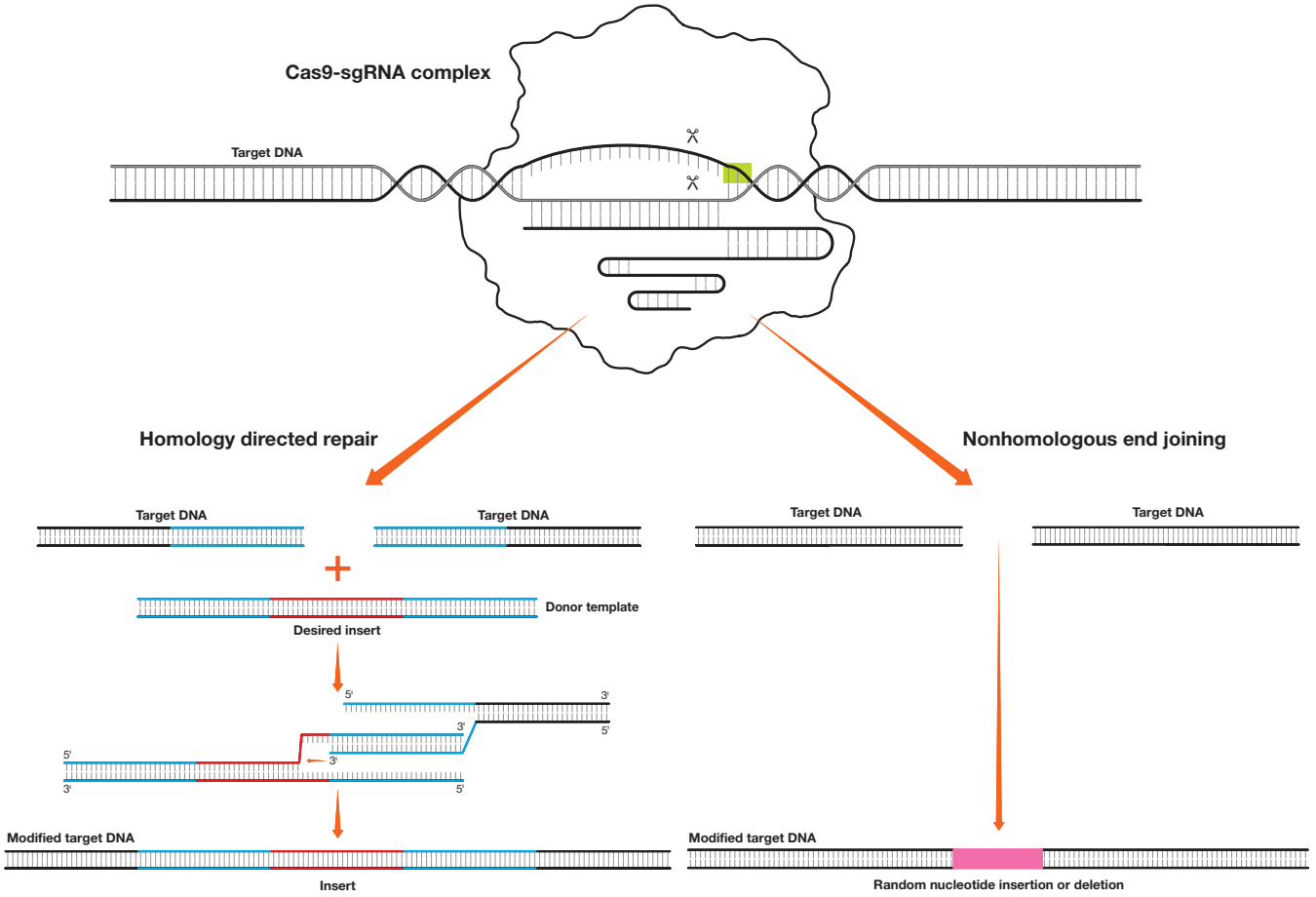


Fig. 4. DNA repair via homology directed repair (HDR) and non-homologous end joining (NHEJ).

The homology arms used in HDR can be hundreds of base pairs or longer. For simplicity, you will simulate basic HDR in this activity using much shorter, 15 bp homology arms. You will design a donor template DNA sequence that could be used to insert a section of DNA into the cut site you created in DNA strip 1.

1. ***Retrieve the two pieces of DNA strip 1. If you have not already used scissors to cut the strip at the cut site, do so now.***
2. ***Cut out the donor template DNA strip and one blank DNA strip.***
3. ***The shaded region simulates a nonspecific DNA insertion sequence. In the empty boxes on either side of the shaded region and on both strands, write the 15 bp sequences that match the nucleotide sequences on either side of the cut site of DNA strip 1. These 15 bp sequences are your homology arms.***
4. ***You now have a complete donor template DNA.***
5. ***Use scissors to cut the excess ends of the donor template strip.***
6. ***Place the pieces of DNA strip 1 directly on top of the donor template strip nucleotide sequences so that the homology arms are aligned and only the insertion sequence is visible. Tape the pieces together.***
7. ***You now have an edited piece of DNA.***

Use the blank DNA strip along with the rest of the paper model pieces to design donor template sequences with 15 bp homology arms that will induce each of the following changes to DNA strip 1 using sgRNA 1. You will need to include any necessary insertion sequences as well as homology arm sequences. Write the final sequences in the table below. Underline the homology arm regions.

Desired Change	Donor Template DNA Sequence
Cause a frameshift	
Insert an EcoRI restriction site	

Focus questions

A. Describe two possible advantages of using HDR over using NHEJ in a gene editing experiment.

B. Describe two possible advantages of using NHEJ over using HDR in a gene editing experiment.

C. Explain how CRISPR-Cas9 together with HDR could be used to change a single nucleotide, for instance changing a T to an A.

D. In addition to inserting or exchanging sequences, it is possible to remove short sequences near a cut site using HDR. Think of and describe an idea for how the donor template DNA sequence could be designed to cause such a removal. Use external resources about HDR as needed.

Activity 2

lacZ CRISPR Gene Editing Laboratory

In this activity you will use CRISPR-Cas9 to cut the bacterial chromosome DNA within the *lacZ* gene, which codes for the enzyme β-galactosidase (β-gal). The *lacZ* gene is part of the *lac* operon, a collection of genes that allows bacteria to use lactose, a milk sugar, as a food source. Conveniently, β-gal also breaks down the colorless compound X-gal into two pieces, one of which is deep blue. If β-gal is expressed by bacteria in the presence of X-gal, they will turn blue. For decades, molecular biologists have used the *lacZ* gene as a target site for inserting DNA sequences because the bacterial colony color indicates whether they were successful. You will use this classic blue-white screening technique as a visual indicator of whether you have successfully edited the *lacZ* gene.

Background

Gene editing

Gene editing involves two steps: cutting double-strand DNA at a desired location and then directing DNA repair to produce a desired sequence change. When chromosomal DNA in a bacterial cell is cut, the cell will die unless it's able to repair the cut. As you saw in the previous activity, cells can repair double-stranded breaks in DNA in several ways, including:

- **Nonhomologous end joining (NHEJ)** — specific proteins reconnect the ends of the double-stranded break back together. This process may randomly insert or delete one or more bases and can cause mutations that can disrupt gene function or expression
- **Homology directed repair (HDR)** — enzymes patch the break using donor template DNA, which is required for HDR. Researchers design the donor template DNA, which may include a desired sequence flanked on both sides by "homology arms" that match the sequence upstream and downstream of the cut. A complementary DNA strand is created during repair

In this activity, you will use CRISPR-Cas9 to cut bacterial chromosomal DNA at a specific location within the *lacZ* gene. You will then take advantage of the cells' ability to perform HDR to cause a desired change in the *lacZ* sequence. You will do this by providing the cells with large quantities of a donor template DNA, which includes an insert with a stop codon that will disrupt the gene function.

The lacZ *gene and blue-white screening*

A gene in the *lac* operon, *lacZ* encodes an enzyme called β-galactosidase (β-gal), which catalyzes the hydrolysis of the sugar lactose into its component sugars. β-gal can also hydrolyze a sugar analog called X-gal, which produces a blue pigment after it is hydrolyzed. Bacteria expressing functional β-gal turn blue when they are grown in the presence of X-gal as shown in Figure 5.

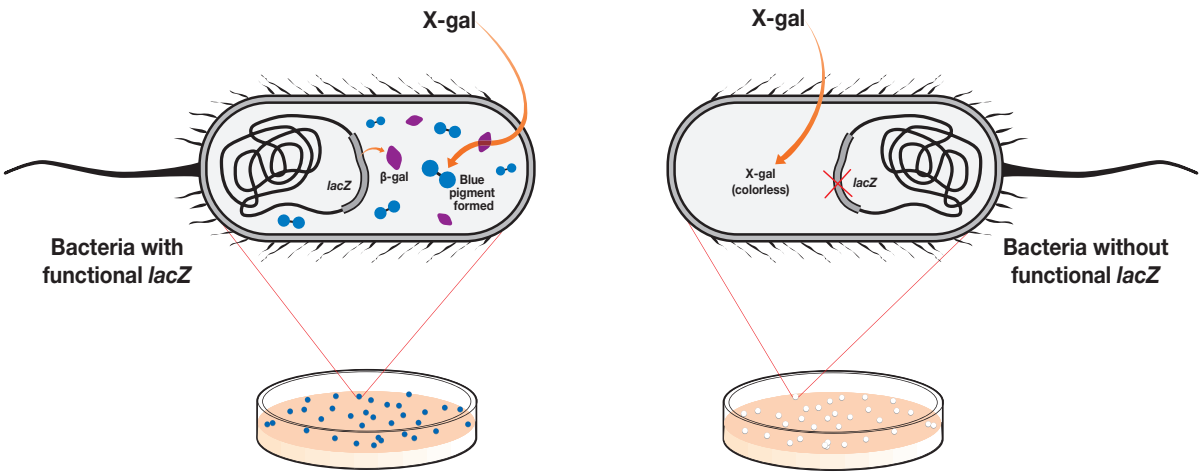


Fig. 5. The function of *lacZ* in blue-white screening.

In nature, lactose induces the expression of the *lac* operon. But because the *lac* operon allows bacteria to use lactose itself as a food source, they consume it, which then stops expression. Therefore to induce continuous expression of *lacZ* scientists use a nonhydrolyzable lactose analog called IPTG in the growth medium to induce β-gal expression.

E. coli *bacteria*

The bacterial strain you will be given at the start of this experiment, *E. coli* HB101-pBRKan, naturally has a functional *lacZ* gene. This particular strain has also been engineered to express Cas9, and it has a plasmid that carries the genes that enable HDR. In these bacteria, expression of the HDR DNA repair system is controlled by an arabinose-inducible promoter; when the bacteria are exposed to arabinose, they express, or "turn on," the HDR DNA repair machinery. Only then can the bacterial cells can use donor template DNA to repair double-strand breaks. Like many lab strains of *E. coli*, the bacteria are modified so that they cannot perform NHEJ. This is for safety reasons.

The cells that have been exposed to arabinose will retain the enzymes needed for HDR even if they are transferred to a plate with no arabinose. Their daughter cells, however, will not produce HDR enzymes unless they are exposed to arabinose.

Plasmids

The bacteria do not normally produce the sgRNA and donor template DNA required to edit the *lacZ* gene. You will introduce sgRNA and/or donor template DNA by transforming bacteria with one of two plasmids:

- pLZDonor — (control) includes a donor template DNA sequence that will be used by the HDR machinery to fix double-stranded DNA breaks. The donor template DNA includes an insert sequence, which will be inserted into the *lacZ* gene and impair its function
- pLZDonorGuide — includes both the donor template DNA sequence from pLZDonor and a sequence that codes for the sgRNA. Once transcribed, the sgRNA will direct Cas9 where to cut *lacZ*

Both plasmids also carry a gene that confers resistance to the antibiotic spectinomycin (SPT).

Part 1. Answer Pre-Laboratory Questions

Table 2. Starter plate conditions.

Starter Plate	Plate Additives	Bacterial Colony Color	Cas9	DNA Repair System	sgRNA	Donor Template DNA
IX	IPTG, X-gal	Blue	+	OFF	–	–
IX/ARA	IPTG, X-gal, arabinose	Blue	+	ON	–	–

A. Using evidence from Table 2, explain in complete sentences why the bacterial colonies on the starter plates are blue.

B. If the bacteria on the starter plates did NOT have a functional* lacZ *gene, what color would you expect the colonies to be?

C. Explain how the differences between the IX and IX/ARA starter plates may influence gene editing in the laboratory activity.

Table 3 lists the four experimental samples (A, B, C, and D) that you will be working with as well the conditions under which they will be grown. During the activity, each sample will be transformed with the plasmids indicated in the Plasmids column.

Table 3. Experimental samples.

Sample	Bacteria Source	Plasmids	Cas9	DNA Repair System	sgRNA	Donor Template DNA	Predicted *lacZ* Change
A	**IX**	pLZDonor	+	OFF			
B	**IX**	pLZDonorGuide	+	OFF			
C	**IX/ARA**	pLZDonor	+	ON			
D	**IX/ARA**	pLZDonorGuide	+	ON			

D. Based on the plasmid that will be added to each sample, fill in the sgRNA and Donor DNA columns with "+" or "–" to indicate which components those bacteria will have.

E. Predict any changes that may occur in the* lacZ *gene during the laboratory activity for each sample. Record your answers in Table 3.

Following transformation, each sample will be spread on LB agar plates that contain additives and incubated to allow colony formation.

Table 4. Bacterial plate cultures.

Plate	Plate Additives	Growth Expected? (Yes/No)	Color of Colonies (If Growth)
A	IPTG, X-gal, spectinomycin		
B	IPTG, X-gal, spectinomycin		
C	IPTG, X-gal, spectinomycin		
D	IPTG, X-gal, spectinomycin		

F. Based on your answers to the previous questions, fill in Table 4 with your predictions of whether there will be bacterial growth on each plate.

Part 2. Conduct Gene Editing

Student workstation

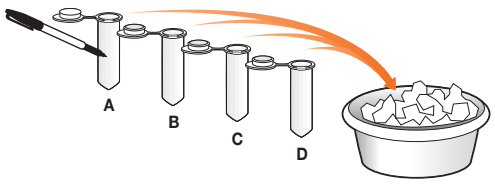
Materials	Quantity
Fresh *E. coli* IPTG/X-gal (**IX**) LB starter plate	1
Fresh *E. coli* IPTG/X-gal/ARA (**IX/ARA**) LB starter plate	1
IPTG/X-gal/spectinomycin (**IX/SPT**) LB plates	4
LB nutrient broth (**LB**)	1.2 ml
Transformation solution (**TS**) on ice	1.5 ml
pLZDonor plasmid (**pD**), 80 ng/µl	25 µl
pLZDonorGuide plasmid (**pDG**), 80 ng/µl	25 µl
100–1,000 µl adjustable-volume micropipet and tips (recommended)	1
20–200 µl adjustable-volume micropipet and tips	1
2–20 µl adjustable-volume micropipet and tips	1
Micro test tube, 2.0 ml	4
Yellow inoculating loop	8
Ice bath with crushed ice	1
Permanent marking pen	1
Foam float (if using water bath)	1
Tube rack (recommended)	1
Waste cup	1

Common workstation

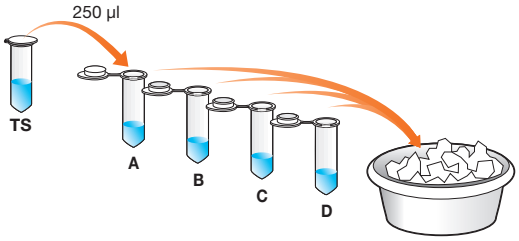
Materials	Quantity
Water bath or dry bath (holes filled with water) at 60°C	1
Incubator oven or shaking incubator with dish shelf at 37°C (recommended)	1
Lab tape	

Protocol

1. ***Label four 2.0 ml microcentrifuge tubes A–D and place on ice.***

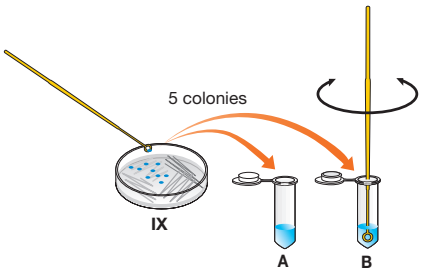


2. ***Add 250 µl ice cold transformation solution (TS) to each tube. Place back on ice.***



3. ***Using a new inoculation loop, pick five colonies from the IPTG/X-gal (IX) plate.***

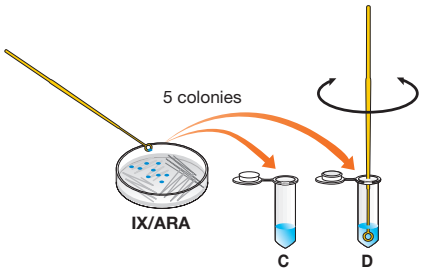
   Swirl the loop in tube A for at least 1 min until all the bacteria are dispersed in the solution. No bacteria should remain on the loop. Immediately place tube back on ice.



4. ***Repeat step 3 for tube B with a new loop.***

5. ***Using a new loop, pick five colonies from the IPTG/X-gal/Ara (IX/ARA) plate.***

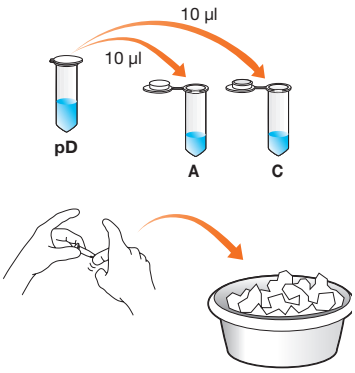
   Swirl the loop in tube C for at least 1 min until all the bacteria are dispersed in the solution. No bacteria should remain on the loop. Immediately place tube back on ice.



6. ***Repeat step 5 for tube D with a new loop.***

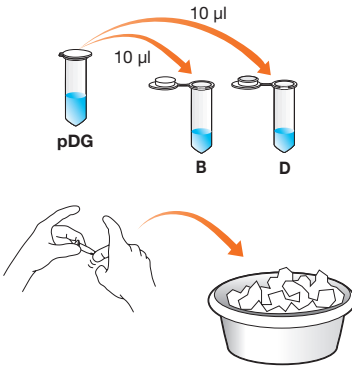
7. ***Using a new pipet tip, add 10 µl pLZDonor (pD) plasmid to tube A. Close the tube, flick three times to mix, and place on ice.***

   Using another new pipet tip, repeat with tube C.



8. ***Using a new pipet tip, add 10 µl pLZDonorGuide (pDG) plasmid to tube B. Close the tube, flick three times to mix, and place on ice.***

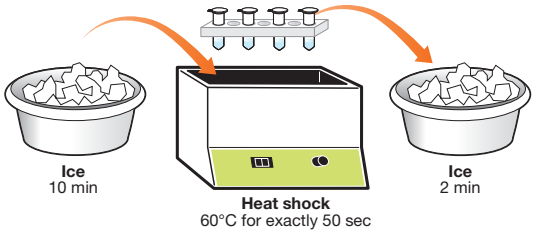
   Using another new pipet tip, repeat with tube D.



9. ***Incubate on ice for at least 10 min.***

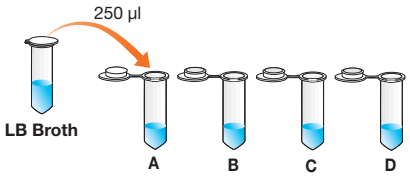
10. ***Bring tubes <u>on ice</u> to the water bath or dry bath.***

    Heat shock at 60˚C for <u>exactly 50 sec.</u> Be sure the bottoms of the tubes contact the water.



11. ***Immediately return the tubes to ice for 2 min. Then transfer to a tube rack.***

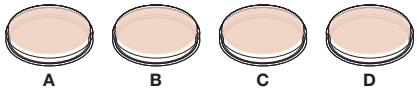
12. ***Using a new pipet tip, add 250 µl LB nutrient broth to each tube. Close each tube and gently flick three times to mix. Leave at room temperature for 20 min to overnight.***



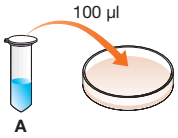
STOP

***Stop.** Ask your instructor whether to proceed now or tomorrow.*

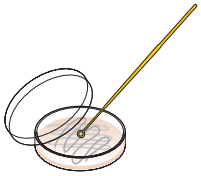
13. ***Near the edges, label the bottoms of four IX/SPT plates A–D. Add your initials and date.***



14. ***Gently flick tube A to resuspend the bacteria. Using a new pipet tip, transfer 100 µl of sample A onto plate A.***

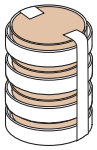


15. ***Using a new inoculation loop, spread the liquid evenly on plate A. Rotate the plate several times in the process. Do not pierce or jab the agar surface.***



16. ***Using a new pipet tip and inoculation loop each time, repeat steps 14 and 15 for samples B–D.***

17. ***Cover, stack, tape, and label your plates. Incubate the plates upside-down at 37°C for 24 hr or at room temperature for 2–3 days.***



18. ***After incubation, check your plates for color development. If blue and white colonies are indistinguishable, refrigerate your plates at 4°C for 1–5 days until the color difference is easily distinguishable.***

Counting colonies and analyzing results

Count the blue and white colonies on your plates and record the numbers in Table 5. Use a permanent marker to mark a dot on the bottom of the plate under each colony as you count it. If there are too many colonies on a plate to count, divide your plate into quadrants and count colonies using steps 19 and 20.

19. ***On the bottom of each plate, use a ruler and a permanent marker to divide the plate into equal quadrants.***

20. Count the blue and white colonies in one quadrant. Use a permanent marker to mark a dot on the bottom of the plate under each colony as you count it. Multiply the number of colonies you counted in one quadrant by four and record your data in Table 5.

Repeat for each plate.

Important!

Before continuing, ask your instructor whether to store your plates for later use in genotyping analysis.

Table 5. Colony counting data.

Plate	# Blue Colonies	# White Colonies	Total # of Colonies	Percentage of White Colonies (# White/# Total)	Comparison with Prediction
A					
B					
C					
D					

21. Calculate the total number of colonies for each plate and record the results in Table 5.

22. For each plate, calculate the percentage of total colonies that are white.

23. Compare your predictions from Table 4 with your results. Record and describe agreements or differences in Table 5. For each difference, provide an explanation.

Part 3. Answer Post-Laboratory Questions

A. Explain how colony color can be used as evidence of the state of the lacZ *gene in the bacteria.*

B. Which plates show evidence of the lacZ *gene having been cut by Cas9?*

C. Of the plates that show evidence of the lacZ *gene having been cut, which also show evidence of the DNA cut having been repaired? Note that repairing DNA does not mean repairing the function of a gene.*

D. What happens to a bacterium if a double-strand DNA break is not repaired?

E. One of your plates may have few if any colonies on it. Write a claim supported by evidence from your results about why colonies did not grow. Include reasoning for why your evidence supports your claim.

F. If you have any unexpected results, list them here and provide a hypothesis for how they occurred.

G. Describe at least two other experiments that could be done to verify that chromosomal gene editing occurred in the bacteria.

Capstone Activity

Identification and Bioinformatics Analysis of Cas9 Target Sites

The ability of the CRISPR-Cas9 system to accurately and permanently edit genomes has major implications for the treatment of diseases. Some diseases, such as coronary artery disease, sickle cell disease, and cystic fibrosis, are caused by genetic mutations. A CRISPR-based therapy that can edit the genomic DNA in cells may be able to correct those mutations.

Though this type of therapy is promising, it is not as clear-cut as it may seem on the surface. As with any therapy, there are risks involved that must be analyzed and understood before testing in humans. For example, off-target effects, where a gene or DNA sequence other than the intended target is edited, can have dire effects on an organism. These types of risk can never be completely eliminated, but their probabilities and the conditions in which they may occur must be evaluated.

One of the first steps in designing a CRISPR-based therapy is identifying a gene-editing strategy and selecting a Cas9 target site to cut. In this activity, you will research the genetic basis for a disease and explore a CRISPR-based gene-editing strategy: replacing, inserting, or deleting a sequence. Then you will identify potential Cas9 target sites and use the basic local alignment search tool (BLAST) from the National Center for Bioinformatics Information (NCBI, part of the National Institutes of Health, NIH) to search for similar sequences in the human genome. Using these data, you will analyze your potential Cas9 target sites for risk of off-target effects to identify the most promising candidate for a CRISPR-based therapy.

Part 1. Identify and Catalog Target Sequences

1. ***Read the background information about the disease you are investigating.***
   Discuss with your group and answer the disease-specific reading questions.
2. ***Scan the provided DNA sequence and identify all potential Cas9 target sequences.***
   Consider the following:
   - A Cas9 target site includes a 20-nucleotide protospacer sequence followed downstream by an appropriate PAM sequence (5'-NGG) in the 5' to 3' direction. Therefore, a target sequence is 23 nucleotides long
   - A search for PAM sequences first may speed up the process
   - Target sequences can be found on either DNA strand, but always in the 5' to 3' direction
3. ***Select 2–4 candidate target sequences to investigate.***
   Record each sequence in the table below, using the following naming convention: Gene name abbreviation-your initials-#. For example, GENE9-TRP-1.

Target Sequence Name	Position # of First Nucleotide	Position # of Last Nucleotide	23-Nucleotide Target Sequence, 5' to 3'

Part 2. Perform BLAST Search for Off-Target Sequences

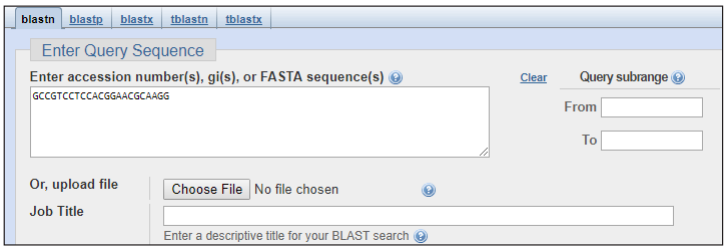
A complete or partial Cas9 target sequence can sometimes be found elsewhere in the human genome, so an sgRNA designed against such a site may guide Cas9 to cut off-target sites. You will use the bioinformatics software BLAST to find genes with sequences that completely or partially match the target sites you selected above.

1. ***Prepare your results table.***
   Use the bioinformatics results table provided in the student guide or recreate it in a spreadsheet program.

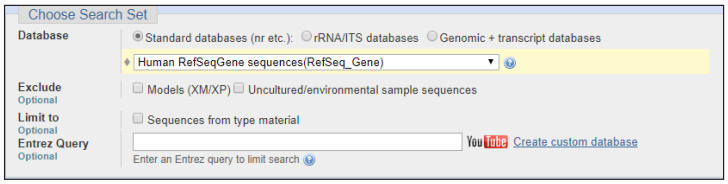
2. ***Perform a BLAST search.***
   The BLAST interface changes frequently. The following instructions and screenshots may deviate slightly from your experience on the site.

   2.1. Visit blast.ncbi.nlm.nih.gov

   2.2. Click Nucleotide BLAST. Copy your target sequence from your table and paste it under Enter Query Sequence.

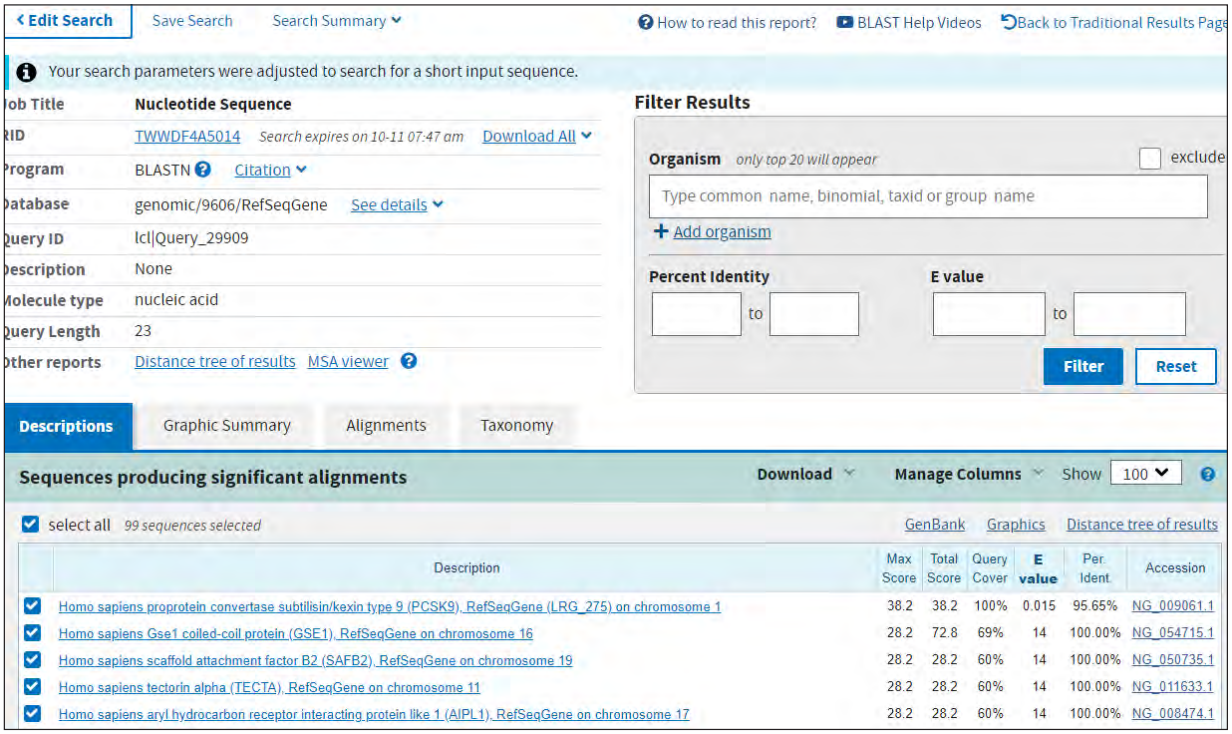


   2.3. Under Choose Search Set > Database, select Human RefSeqGene sequences (RefSeq_Gene).



*2.4. Select **Show results in a new window** and click **BLAST**.*

2.5. The BLAST search may take a few minutes to complete, depending on the server usage volume. When complete, a screen similar to that below appears.

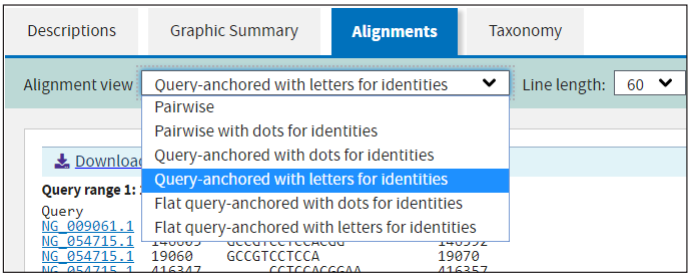


3. Review the BLAST search results.

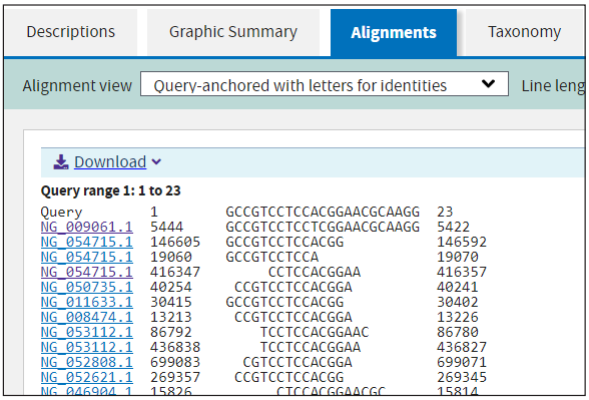
3.1. One of the top results should be an exact match for your query in the gene you are working with. Click the description link and review the information provided. Then return to the BLAST results page.

3.2. Click other links in your results list to orient yourself to the information provided. Then return to the BLAST results page.

*3.3. Select all the sequences (select all), click to open the **Alignments** tab, then select **Alignment view > Query-anchored with letters for identities.***



3.4. Review the sequence information.



a. *Your sequence (Query) appears at the top of the results for easy reference. The remaining rows display nucleotide sequences from genes that match or partially match your query sequence.*
b. *The links at left are the accession numbers for reference genes. Each gene in the database has its own accession number.*
c. *To the right of each accession number is the start position of the alignment sequence, followed by the sequence and the number of the nucleotide at the end position. Most alignment sequences will not match the full length of the query sequence.*

3.5. One of the top results should be an exact match for your query and be located in the gene you are working with. Click the accession number. Ensure the information given for this match is correct (that it matches the gene you are working with). If it does not match exactly, check that your query sequence is correct.

4. Annotate your results.

4.1. In your bioinformatics results table record five alignment results (or as many as possible if there are fewer than five) whose sequences include the PAM sequence 5′-NGG aligned with that of the query sequence, where N is any nucleotide. Use the accession number link to retrieve additional information to fill in the table. There may be multiple results for a single accession number.

4.2. Highlight any exact alignment matches in genes other than your intended target.

4.3. Circle the longest alignment sequence(s) other than your intended target.

5. Repeat steps 2–4 for each of your target sequences.

Focus questions

A. Why might there be multiple results from a single accession number?

B. How might an exact alignment match be an off-target cut site?

C. Does the presence of alignment matches indicate higher or lower risk of off-target effects?

Part 3. Evaluate Candidate Target Sequences

Because the sgRNA sequence determines where Cas9 cuts DNA, it may cut DNA anywhere it encounters a complementary sequence, even in undesired locations. Such unintended cuts are called off-targets. Experiments have shown that Cas9 sometimes cuts DNA even when there are mismatches between the sgRNA and DNA sequences. This means DNA sequences that only partially match the sgRNA have the potential to be off-target cut sites, although the PAM sequence is still required. In general, the more mismatches there are between a DNA sequence and the sgRNA, the less likely Cas9 is to cut.

1. ***Develop three criteria that could be used to rank the candidate target sites according to their safety for therapeutic use. Use the alignment results you have collected in your BLAST search as well as the information you considered from the readings and focus questions as you select criteria. For each criterion, explain how it will help you evaluate a potential target site.***

2. ***Use your criteria to evaluate and rank the candidate target sites you selected in Part 1, step 3. Write a claim about which candidate target site is the best option for your therapeutic application, based on your evaluation, and provide reasoning.***

Focus questions

A. If you were to continue evaluating the candidate target sites for use in a therapy, what are two additional pieces of information or experiments that would help you?

B. What health problems could arise from off-target CRISPR-Cas9 activity?

C. How would you decide whether the risk of off-target activity for a CRISPR-Cas9 therapy is low enough to be considered safe?

D. Should off-target effects be considered for nontherapeutic CRISPR experiments in the laboratory? Explain why or why not.

E. Do you think there should be differences between how off-target risk is evaluated for CRISPR-based therapies and for laboratory CRISPR experiments?

Coronary Artery Disease

Background

Cardiovascular disease is the leading cause of death worldwide, claiming over 17 million lives annually. One type of cardiovascular disease, coronary artery disease (CAD), in which blood vessels near the heart become narrowed due to plaque buildup, claims nearly 8 million lives each year. Lowering levels of low density lipoprotein (LDL) cholesterol has been shown to effectively reduce risk of CAD. LDL receptors (LDLR) in the liver clear LDL from blood plasma. However, levels of LDLR are themselves reduced by proprotein convertase subtilisin/kexin type 9 (PCSK9), a serine protease that binds and degrades the receptors (Figure 6). People with mutations in the *PCSK9* gene commonly have lower levels of LDL cholesterol likely because they have higher levels of the LDL-clearing receptors.

Gene-Editing Therapy Strategy

A goal of gene-editing therapy may be to reduce or eliminate PCSK9 enzyme function. One strategy is to disrupt the gene by making a cut within exon 1 and allowing nonhomologous end joining (NHEJ) to occur. This strategy would reduce levels of functional PCSK9 enzyme in the liver, which would reduce degradation of the LDL receptors to allow more removal of LDL cholesterol from the bloodstream.

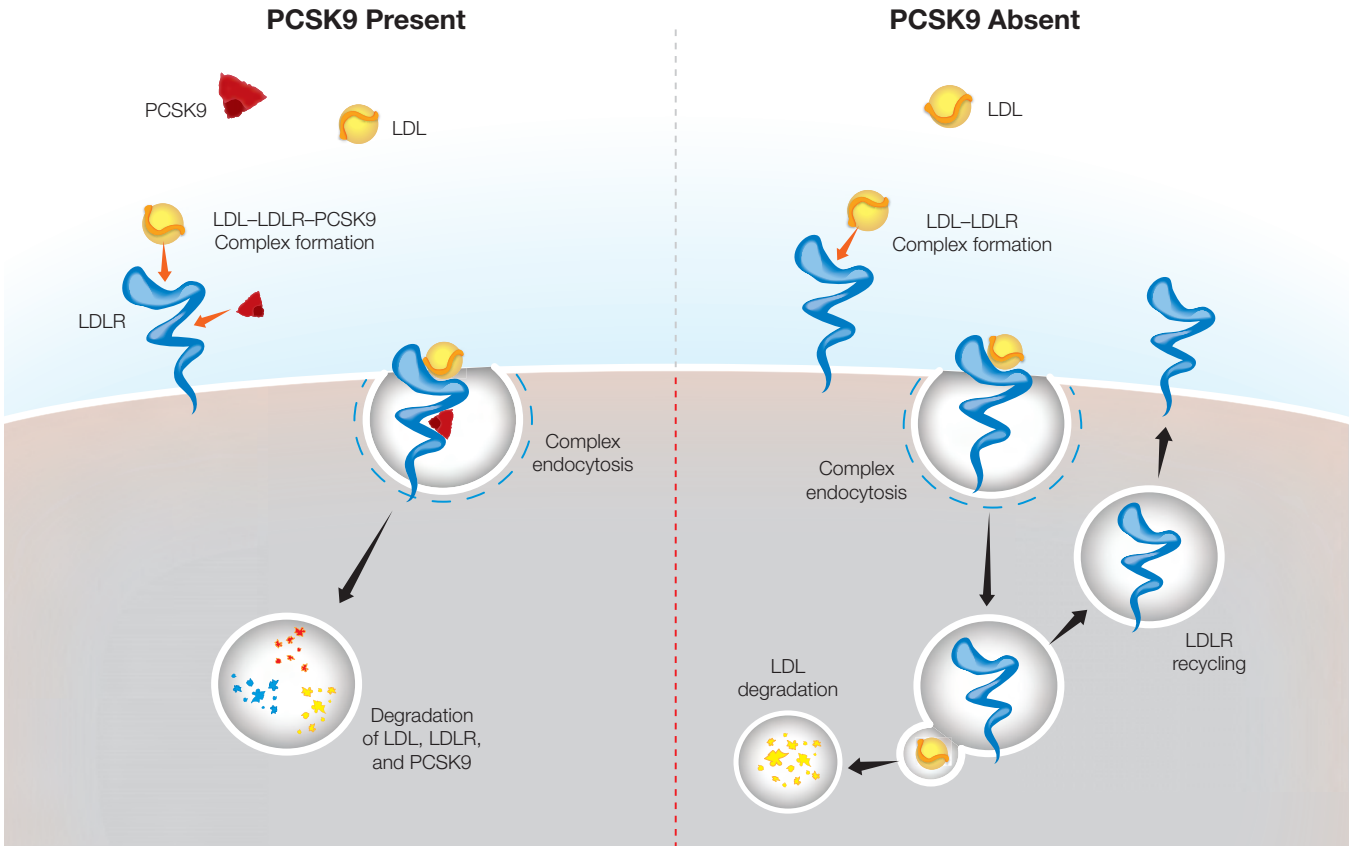


Fig. 6. Degradation of LDL in the presence and absence of PCSK9.

Focus questions

A. Describe how disrupting **PCSK9** *would impact gene expression. Draw a model that illustrates your description.*

B. Describe two potential advantages and two potential disadvantages of administering such a therapy to liver cells only.

C. Is this gene editing strategy an example of replacing, inserting, or deleting a sequence?

PCSK9 *gene sequence information*

Gene Accession Number: ***NG_009061.1***
Gene Reference: ***Homo sapiens proprotein convertase subtilisin/kexin type 9***
Gene Abbreviation: **PCSK9**

The sequence below is an excerpt of *PCSK9* exon 1 nucleotide position 5,387 to 5,446.

```
5387 5'-GGACGAGGACGGCGACTACGAGGAGCTGGTGCTAGCCTTGCGTTCCGAGGAGGACGGCCT-3' 5446
     3'-CCTGCTCCTGCCGCTGATGCTCCTCGACCACGATCGGAACGCAAGGCACCTCCTGCCGGA-5'
```

Sickle Cell Disease

Background

Sickle cell disease is an inherited blood disorder in which a person's red blood cells become sickle-shaped, which increases the risk of blood clots. When we scrape a knee or cut a finger, blood clots form externally to create a scab over the wound and promote healing. When blood clots form internally, however, they can block blood vessels and cause pain or even death. Approximately 100,000 people die of complications from sickle cell disease each year.

Sickle cell disease is caused by a single nucleotide polymorphism (SNP) in the hemoglobin B *(HBB)* gene called rs334. People homozygous for adenine at rs334 produce normal hemoglobin while those homozygous for thymine at rs334 produce sickling hemoglobin and have the disease. Heterozygous individuals do not exhibit symptoms of the disease and have increased resistance to malaria, which increases their evolutionary fitness in regions where malaria is common.

Gene-Editing Therapy Strategy

The goal of gene-editing therapy for sickle cell disease is to allow expression of functional non-sickling hemoglobin. Red blood cells, which carry hemoglobin, are formed from hematopoietic stem cells in bone marrow (Figure 7). A potential gene-therapy strategy is to harvest hematopoietic stem cells from a patient, edit the *HBB* gene in those cells, and reintroduce them to the patient. Using a patient's own cells greatly reduces the chance of rejection by the patient's immune system. CRISPR-based gene editing is used to replace thymine with adenine at rs334 by making a cut near the SNP and introducing the correct sequence using homology-directed recombination (HDR).

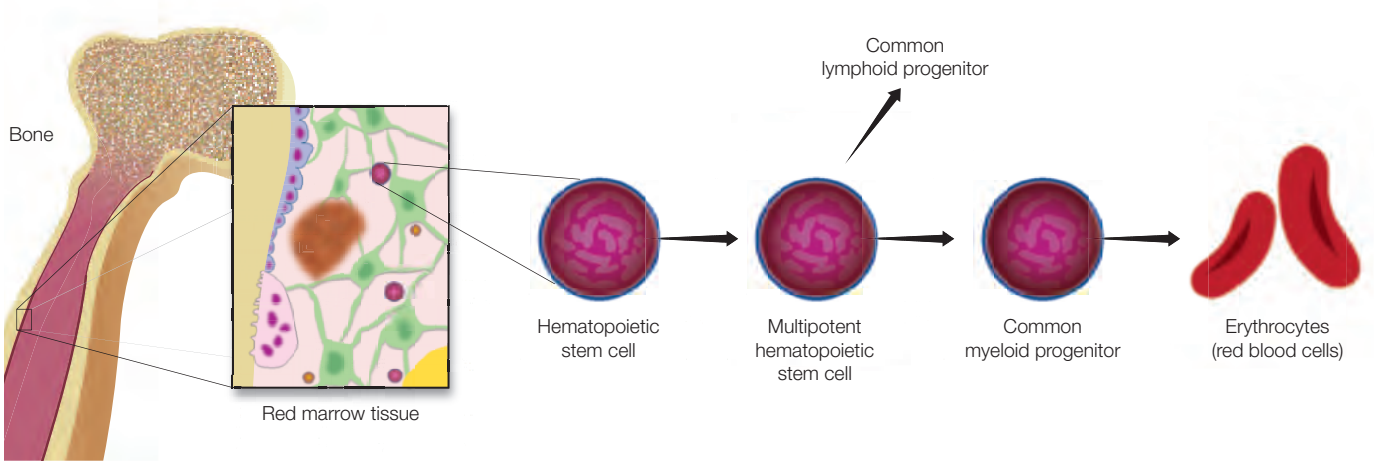


Fig. 7. Differentiation of hematopoietic stem cells into red blood cells.

Focus questions

A. For this gene-editing strategy, why is it useful to edit the DNA of only hematopoietic stem cells?

B. What other cells could be edited to achieve similar results? List two potential advantages and two potential disadvantages of editing these cells instead of hematopoietic cells.

C. Is this gene editing strategy an example of replacing, inserting, or deleting a sequence?

HBB *gene sequence information*

Gene Accession number: ***NG_059281.1***
Gene Reference: ***Homo sapiens hemoglobin subunit beta*** **(HBB),** ***RefSeqGene on chromosome 11***
Gene Abbreviation: **HBB**

The sequence below is an excerpt of *HBB*, nucleotide position 5,053 to 5,106, with rs334 shown bolded and with an asterisk.

```
                          *
5053  5'-GGTGCATCTGACTCCTGTGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGT-3'  5106
      3'-CCACGTAGACTGAGGACACCTCTTCAGACGGCAATGACGGGACACCCCGTTCCA-5'
```

Cystic Fibrosis

Background

Cystic fibrosis (CF) is an autosomal recessive disease that affects the lungs, pancreas, and small intestine. The disease affects about 70,000 individuals worldwide. It causes buildup of viscous mucus in these organs and frequently leads to severe lung infections. If untreated, most CF patients do not live past their 20s. The disease is caused by mutations in the cystic fibrosis transmembrane conductance regulator *(CFTR)* gene on chromosome 7, the most common of which is an in-frame deletion of three base pairs in exon 11 that codes for phenylalanine (F508del). *CFTR* is a protein that transports chloride ions across cell membranes, which is critical to effective clearing of mucus from airways (Figure 8). The F508del mutation impairs the normal production of CFTR.

Gene-Editing Therapy Strategy

The goal of cystic fibrosis gene-editing therapy is to correct the *CFTR* gene in lung epithelial stem cells. A therapeutic drug will likely be given by inhalation to target the lungs and the edit will occur *in vivo*. CRISPR technology would be used to create a cut in exon 11 in the vicinity of the *CFTR* F508del mutation and use CRISPR-mediated homologous recombination to replace the mutation with a healthy version of the gene.

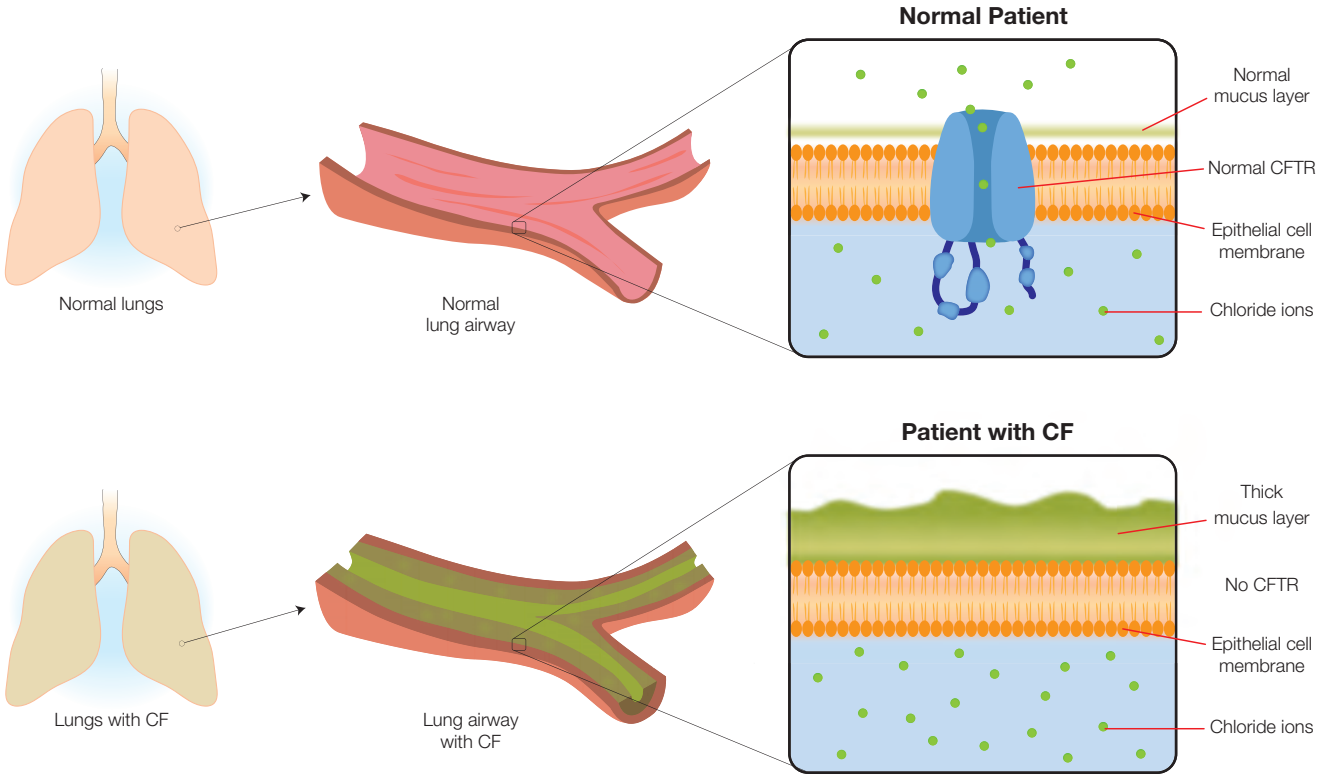


Fig. 8. The mucus layers in lung airways of healthy patients and those with CF.

Focus questions

A. Describe how editing **CFTR** *would impact gene expression. Draw a model that illustrates your description.*

B. What are two potential advantages and two potential disadvantages of administering a cystic fibrosis gene editing therapy by inhalation instead of orally, by injection, or another method?

C. Is this gene editing strategy an example of replacing, inserting, or deleting a sequence?

CFTR *gene sequence information*

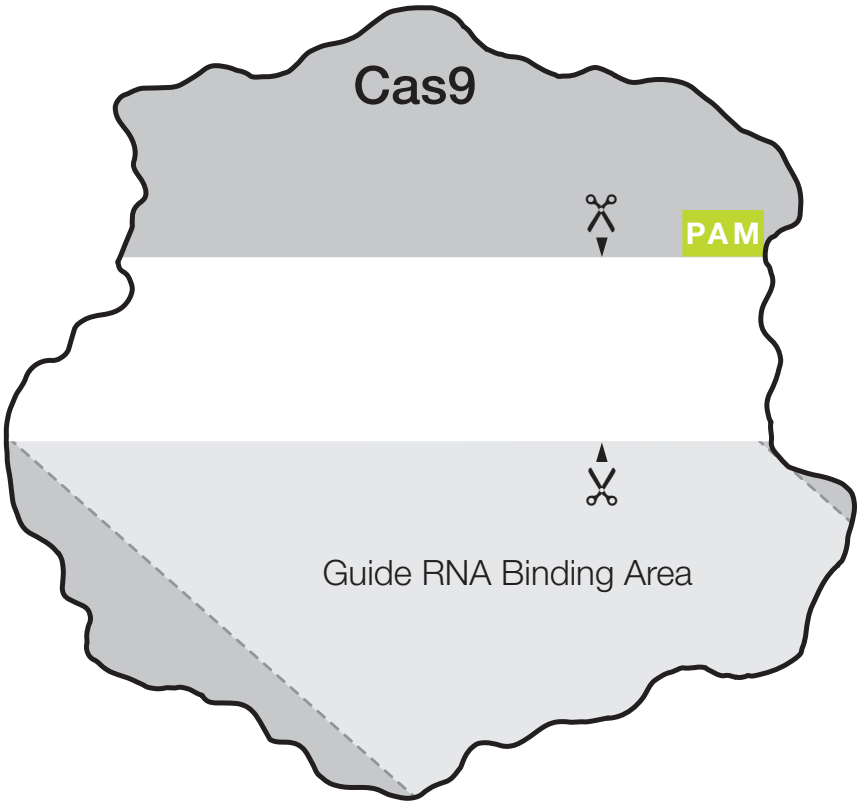
Gene Accession Number: *NG_016465.4*
Gene Reference: *Homo sapiens CF transmembrane conductance regulator* **(CFTR)**
Gene Abbreviation: **CFTR**

The sequence below is an excerpt from exon 11 of CFTR, nucleotide position 98,756 to 98,815, with the three-nucleotide mutation shown bolded and with asterisks.

```
                                                        ***
98756   5'-TTCTGTTCTCAGTTTTCCTGGATTATGCCTGGCACCATTAAAGAAAATATCATCTTTGGT-3'   98815
        3'-AAGACAAGAGTCAAAAGGACCTAATACGGACCGTGGTAATTTCTTTTATAGTAGAAACCA-5'
```

Target Sequence Name	Target Sequence (5′ to 3′)	Gene Accession Number of Result	Gene Name (Abbreviation)	Alignment Sequence (5′ to 3′)	Start Nucleotide #	End Nucleotide #	Length	Alignment Sequence Includes Correct PAM? (Y/N)	Number of Mismatches

3' – AGTCCAGTTTAAGTCTGCCGTTTGCTGACAGGACCGGCATTGGCTGGGTCGGCGGGCAACGTGGTGTCTAC – 5'
3' – GTAGACACCACGTTGCCCGCGACCCAGCCAATGCCGGTCCTGTCAGGCAAACGGCAGACTTAAACTGGACT– 5'
DNA STRIP 2
3' – AGTCCAGTTTAAGTCTGCCGTTTGCTGACAGGACCGGCATTGGCTGGGTCGGCGGGCAACGTGGTGTCTAC – 5'
3' – GTAGACACCACGTTGCCCGCGACCCAGCCAATGCCGGTCCTGTCAGGCAAACGGCAGACTTAAACTGGACT– 5'
DNA STRIP 1
5' – 3'
BLANK DNA
5' – 3'
BLANK DNA
5' – 3'
BLANK DNA
5' – 3'
DONOR TEMPLATE DNA
sgRNA 2
AACGGGCGCUGGGUCGGUUA
sgRNA 1


Cas9
PAM
Guide RNA Binding Area

Glossary

Arabinose-inducible promoter — promoter that occurs naturally in bacterial systems and that is used in many expression plasmids to allow regulation of expression of a target gene: expression is induced in the presence of arabinose and repressed in its absence.

Cas9 — CRISPR-associated protein 9 (Cas9), an endonuclease that forms a double-strand break (cuts) in DNA at a specific site within a larger recognition sequence, or target site. It is involved in the natural defense of certain prokaryotes against DNA viruses, and it is heavily utilized in genetic engineering applications to cut DNA at locations specified by a guide RNA (gRNA).

CRISPR — clustered regularly interspaced palindromic repeats (CRISPR) are sequences in the genomes of some prokaryotes that act as a genomic record of previous viral attack. Along with CRISPR-associated (Cas) proteins, bacteria use the sequences to recognize and disarm future invading viruses. Scientists have adapted this system for genetic engineering purposes.

Donor template DNA — engineered sequence of DNA required for homology-directed repair in CRISPR gene editing applications; may include a desired sequence flanked on both sides by "homology arms" that match the sequence upstream and downstream of the cut.

β-galactosidase — encoded by the *lacZ* gene, this enzyme hydrolyzes galactose-containing carbohydrates, including lactose. Conveniently, it also breaks down the colorless compound *X-gal* into two pieces, one of which goes on to form a deep blue pigment.

Guide RNA (gRNA) — non-coding, short RNA sequence that binds to Cas9 and to complementary target DNA sequences, where Cas9 performs its endonuclease activity to cut the target DNA strand.

Guiding region — part of the CRISPR RNA or crRNA in nature, a typically 20-nucleotide region of sgRNA that is complementary to the target DNA sequence and that defines where Cas9 cuts. Scientists can easily customize this sequence for their own targets.

Homology directed repair (HDR) — DNA repair mechanism in which specific proteins patch a double-strand DNA using donor template DNA.

Isopropyl β-d-1-thiogalactopyranoside (IPTG) — a non-metabolizable analog of lactose, which induces transcription of the lac operon.

lacZ — part of the *lac* operon in *E. coli*, this gene encodes the enzyme β-galactosidase. For decades, molecular biologists have used the *lacZ* gene as a target site for inserting DNA sequences because the resulting bacterial colony color indicates whether insert was successful.

Non-homologous end joining (NHEJ) — DNA repair mechanism in which specific proteins reconnect the ends of a double-strand DNA break. This process may randomly insert or delete one or more bases that can disrupt gene function or expression.

Protospacer — DNA region targeted for cleavage by the CRISPR system.

Protospacer adjacent motif (PAM) — sequence motif immediately adjacent to the protospacer sequence in the Cas9 recognition sequence that is required for Cas9 function. Cas9 recognizes the PAM sequence 5'-NGG where N can be any nucleotide (A, T, C, or G). When Cas9 binds the PAM, it separates the DNA strands of the adjacent sequence to allow binding of the sgRNA. If the sgRNA is complementary to that sequence, Cas9 cuts the DNA.

Scaffold region — called the trans-activating CRISPR RNA or tracrRNA in nature, a region of sgRNA that forms a multi–hairpin loop structure (scaffold) that binds tightly in a crevice of the Cas9 protein. The sequence of this region is typically the same for all sgRNAs.

Single guide RNA (sgRNA) — engineered form of guide RNA that forms a complex with Cas9; ~100 nucleotide fusion of two regions that occur as separate guide RNAs in nature: the guiding region (crRNA) and the scaffold region (tracrRNA).

X-gal — 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside, a compound consisting of galactose linked to a substituted indole. Its hydrolysis by β-galactosidase yields an insoluble blue pigment and can be used in bacterial cultures to indicate the presence of active β-galactosidase.

Bio-Rad Laboratories, Inc.

Life Science Group

Website *bio-rad.com* ***USA*** *1 800 424 6723* ***Australia*** *61 2 9914 2800* ***Austria*** *00 800 00 24 67 23* ***Belgium*** *00 800 00 24 67 23* ***Brazil*** *4003 0399*
Canada *1 905 364 3435* ***China*** *86 21 6169 8500* ***Czech Republic*** *00 800 00 24 67 23* ***Denmark*** *00 800 00 24 67 23* ***Finland*** *00 800 00 24 67 23*
France *00 800 00 24 67 23* ***Germany*** *00 800 00 24 67 23* ***Hong Kong*** *852 2789 3300* ***Hungary*** *00 800 00 24 67 23* ***India*** *91 124 4029300* ***Israel*** *0 3 9636050*
Italy *00 800 00 24 67 23* ***Japan*** *81 3 6361 7000* ***Korea*** *82 2 3473 4460* ***Luxembourg*** *00 800 00 24 67 23* ***Mexico*** *52 555 488 7670*
The Netherlands *00 800 00 24 67 23* ***New Zealand*** *64 9 415 2280* ***Norway*** *00 800 00 24 67 23* ***Poland*** *00 800 00 24 67 23* ***Portugal*** *00 800 00 24 67 23*
Russian Federation *00 800 00 24 67 23* ***Singapore*** *65 6415 3188* ***South Africa*** *00 800 00 24 67 23* ***Spain*** *00 800 00 24 67 23* ***Sweden*** *00 800 00 24 67 23*
Switzerland *00 800 00 24 67 23* ***Taiwan*** *886 2 2578 7189* ***Thailand*** *66 2 651 8311* ***United Arab Emirates*** *36 1 459 6150* ***United Kingdom*** *00 800 00 24 67 23*

10000122376 Ver D

22-0526 0822 Sig 0122