

Third Generation Genetic Scissors: **CRISPR**



software
education module

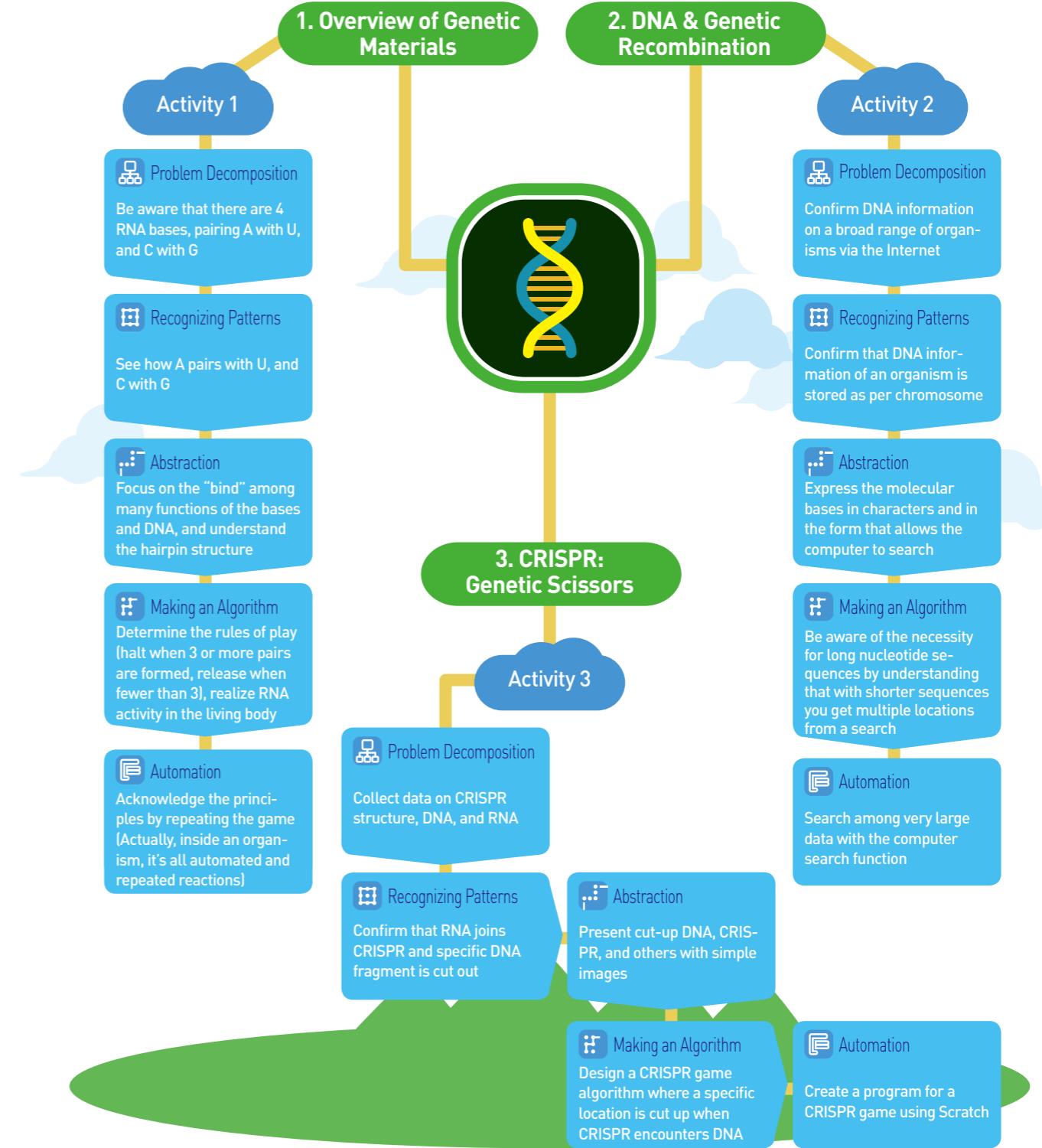
Computational Thinking Map

How to Use This Program

Software is changing the world. The programs installed in computers and apps that make it more convenient to use smartphones are all software. Software is in every part of our lives, so it is difficult to find areas where we are not affected by software.

The state-of-the-art science and technology that we see in the news is also helped by software. In turn, progress in math, science and technology advances software further. As such, math, science and technology are closely related with and cannot be separated from software.

These module series were created through collaboration between experts in related fields and software education, and its suitability for classrooms has been verified. As students follow teachers' direction through each module, they will be able to better understand the world that has been changed by software.

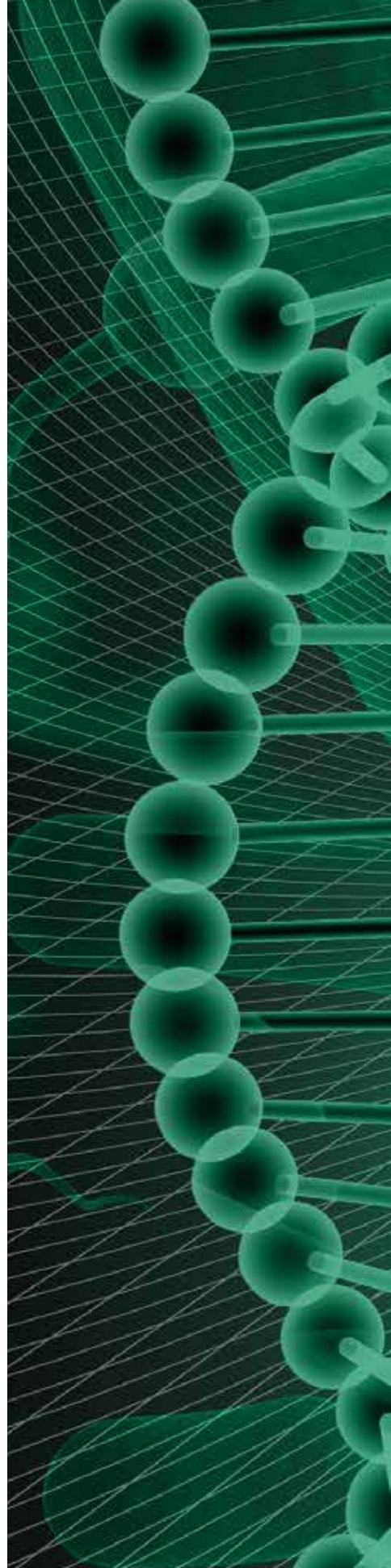


software

education module

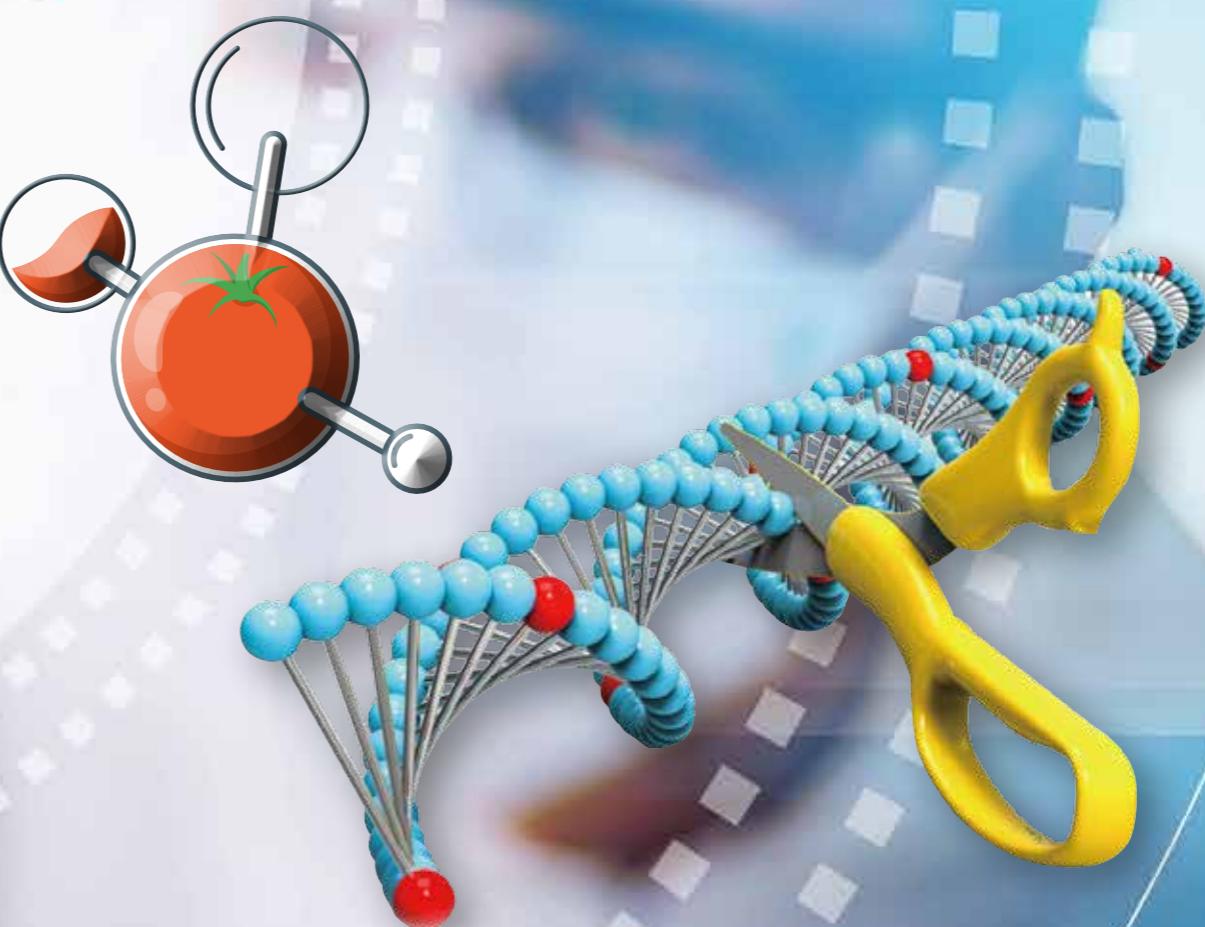
CONTENTS

Intro	The Era of Genetic Engineering	02
PART1	Genetic Materials: An Overview	08
	Activity1 Functions of the Bases in RNA	12
PART2	DNA Coding & Gene Recombination	14
	Activity2 Find a Specific Nucleotide Sequence!	20
PART3	CRISPR: Genetic Scissors	22
	Activity3 Make Your Own CRISPR Game	26
Career Exploration	Bioinformatics Specialist	32
Maker Activity	CRISPR: Genetic Scissors	34



The Era of Genetic Engineering

We are now living in the age of "CRISPR," which refers to genetic scissors used to cut specific genes precisely. We can now edit genes with this tool - virtually any genes as necessary. Suddenly, the genetic manipulation that has proven rather elusive is now right in front of us. In this era of CRISPR and genetic engineering, what will our response be? Before answering this question, let's look at what you need to know about CRISPR.



THE AGE OF GENETIC ENGINEERING AT OUR DOORSTEP

In Florida in November 2016, where one of the fiercest campaigns took place leading up to the US presidential election, residents had to cast their vote on another issue. A referendum was being held in the island community of Key Haven, Florida, on a proposal to release genetically modified (GM) mosquitoes.

Earlier that year, in August, the FDA approved such a release. These genetically altered mosquitoes, developed by British biotech firm Oxitec, were all male, and capable of mating with females in the wild. Larvae from the mating would not grow into adulthood due to toxins from the male GM mosquitoes. Repeated practice of this process would decrease the number of mosquitoes in the targeted region. Experiments in countries such as Brazil and Panama have shown that the number of mosquitoes decreased by at least 80-90% over the course of a year.

Despite the FDA's approval, Florida was unable to conduct the experiment due to opposition of some of the residents and environmental activist groups. The state government decided to hold a referendum to clearly identify public opinion on the issue. The result: 65% of the residents of Key Haven were against the release. However, 58% of the Monroe County population, to which the community of Key Haven belongs, were in favor of the proposal, and the experiment was finally given the green light.

Yet, many Key Haven residents remain worried about the potential for GM mosquitoes to disturb the balance of the ecosystem. According to Oxitec, it is difficult to tell male GM mosquitoes from female mosquitoes, and thus some 0.2% of the total released were female. Unlike their male counterparts, female mosquitoes do the biting to nourish themselves on blood - hence the worry that people may ultimately be harmed. The FDA, however, strongly supports the experiment, arguing that no evidence has been found that any DNA exchange occurs from a mosquito bite.

Genetically-Modified Food - Is It Reliable?

In November 2015, the FDA approved GM salmon to be marketed for human consumption, marking the first time GM animals have been approved for food. Many experiments and verification processes have been carried out to reassure the public of the



AquaAdvantage salmon (background) that has been genetically engineered to grow more rapidly than its non-GM salmon counterpart (foreground) ©AquaBounty Technologies

safety of GM food; yet consumer responses are discouraging. Some large retail distributor companies have flatly stated that they will not sell GM salmon.

Why do people fear GM foods? Are they really that bad?

Currently, the most popular technique for producing such food is with germs (bacteria). By inserting new genes carrying germs into an animal or a plant at will, you can increase production yield or obtain other targeted results. Scientists, however, sometimes do not know exactly what kind of genes are implanted. Thus, GM foods have been a thorny issue due to the risk of unexpected side effects. So far, there has been no scientific evidence to support the claims that GM foods are particularly harmful to humans.

Moreover, the technique to create genetically-modified mosquitoes or salmon is different from the technique used to produce GM crops. For these animals, a special, sophisticated genetic manipulation method is employed to alter several gene sequences* or insert new genes. There are no unidentified genes, and genes that have been confirmed to be safe for humans are most commonly used. For this reason, scientists insist that, instead of the term "genetic manipulation," "gene editing" would be more appropriate as it implies precise modification of genes.

Some skeptics argue that using "gene editing" or "genetic manipulation" does not make a difference

in that the genes are, after all, altered in a reckless manner. But scientists say that there is not much to fear in the products of these techniques.

Humans have some 3 billion DNA base pairs, which means that in order to create a new cell (for example, a sperm or an egg), it is necessary to copy as many as 3 billion DNA base pairs into it. Cells replicate DNA very precisely, but each cell division produces over 130 mutations*. Most mutations, however, do not cause problems as they do not occur in areas where proteins are made. In a way, the logic is that altering one or two genes to produce GM foods is actually not a big deal.

In August 2016, 111 Nobel laureates called on environmental advocacy groups that claim GM crops are dangerous to cease their campaigning against them. These laureates said that such claims lack evidence, arguing that in order to resolve food issues, the development of genetically-modified organisms (GMO) should be encouraged.

So far, genetic engineering is, scientifically, a safe operation. If, after accumulating sufficient experimental data, we edit genes precisely, safety concerns that exist now will largely disappear.

* **gene sequence:** the order of the four bases in a strand of DNA

* **mutation:** a sudden change in DNA that can be transferred to offspring

WITH CRISPR, ANYONE CAN BE A GENE EDITOR

It has been a longstanding dream of most scientists to develop technologies that can manipulate genes precisely. Most notable among these techniques is the use of “gene scissors” that will snip only targeted genes.

The first gene scissors were developed in the 1970s when modern biology was just beginning. Germs protect themselves from foreign genes with “restriction enzymes.” Scientists have adopted these substances to cut out specific genes. Inaccuracy in the manipulation of restriction enzymes, however, was an obstacle to application on higher organisms such as lab rats or humans.

In the 2000s, more advanced gene scissors were introduced. But the process of genetic manipulation using the tool was complicated enough to make it extremely difficult for most people unless they were experts in gene scissors, resulting in slow research efforts. Worse still, with this tool you needed to insert thousands of new DNA sequences to repair just a single gene - which was largely seen as an incredible amount of effort for a tiny outcome.

CRISPR: The Result of 30 years of Collaborative Effort in the Scientific Community

Against this backdrop, a breakthrough came in 1987 with some new gene scissors known today as CRISPR. When a Japanese scientist was looking at the gene sequences of *E. coli*, he found a repeated

palindrome structure (see p. 11). In 1994, it was found that the repeated sequence of palindrome structures occurs in other bacteria as well. Scientists named this mysterious sequence “clustered regularly interspaced short palindromic repeats”, or CRISPR.

In 2005, the sequence was clearly identified in Denmark, where researchers at the dairy producer, Danisco, were struggling with viruses that attack lactic acid bacteria essential in fermenting of dairy products. For some reason, some of the lactic acid bacteria survived despite repeated virus infections. The researchers checked the bacteria genes and found that 21 viral DNA base pairs were stored in CRISPR. Danisco’s scientists also discovered that the lactic acid bacteria store the viral DNA base sequences in their CRISPR and use the sequences to counter viral attacks - similar to identifying a criminal suspect with a composite sketch.

In 2012, Prof. Jennifer Doudna at University of California (Berkeley) in the United States and Prof. Emmanuelle Charpentier at Germany’s Max Planck Institute announced, for the first time, a technique to snip a targeted gene with CRISPR. The two scientists first identified how CRISPR works. When attacked by a viral infiltration, germs repel the viruses and then store 21 viral DNA base pairs in their CRISPR. When attacked again, the 21 sequences act as a guide to summon Cas9 proteins (CRISPR associated protein 9) to the scene. The

Cas9 protein snips out the viral DNA held by the guide sequence, thereby neutralizing the virus.

The joint research team of Professors Doudna and Charpentier succeeded in cutting out the targeted gene by replacing the 21 guide sequences with other genes (for example, by retaining cancer-inducing genes as a guide sequence). The process is so simple and convenient that with only 21 guide sequences and Cas9 protein you can snip any specifically-targeted gene.

Could Scientists Bring Back the Wooly Mammoth with CRISPR?

Scientists are engaged in a range of attempts to create new lab animals or highly efficient GM foods with CRISPR crops.

Prof. George Church at Harvard University has declared that he will bring back, with CRISPR, the wooly mammoth, which went extinct tens of thousands of years ago. Here is his bizarre-sounding plan: First, obtain a well-preserved mammoth carcass, and extract DNA from it. While such DNA cannot be readily used due to the damage it has suffered, the DNA sequence would be read and compared to that of the elephant so genetic differences could be identified.

With CRISPR, the genetic characteristics of the mammoth, such as its thick coat of long hair that enabled it to survive the ice age, would be inserted



Prof. George Church has announced that he is working on a “mammoth revival project” to extract genes from the carcass of mammoths frozen in the Arctic glaciers and insert them into the genes of an elephant. © Geneticist and Mammoth

into the genes of an elephant and used to replicate an animal that resembles the mammoth. Certainly, this would not be an authentic mammoth, and in addition to CRISPR, many other technologies would be needed for success. Nonetheless, it can be admitted that, thanks to CRISPR, scientists are now at least able to attempt such a task.

Gene therapeutics draw more attention than a potential revival of the mammoth, for the possibilities of treating disease through manipulation of human genes. Researchers at Sichuan University in China will begin clinical trials* in 2017, using CRISPR, to treat patients with lung cancer. After extracting white blood cells from persons with lung cancer, the researchers plan to manipulate the genes so that these cells will attack the cancer cells more effectively. It will take a long time, but it is much expected that CRISPR will play an essential role in eradicating intractable diseases such as AIDS and hemophilia.

* **clinical trial:** controlled testing on animal or human subjects to assess the safety of new investigational drugs, etc.

PART 1

Recommended target
Middle school free semester

Relevant subjects
Science and human civilization,
3rd grade science

Genetic Materials: An Overview

* gene

Genes are the basic unit of genetic information, which determines a person's or animal's characteristics inherited from their parents. As an analogy to the computer, genes are the information coded in a hard disc. It was previously believed that one gene controlled one characteristic, but it turns out to be much more complicated than that, and in fact, a broad range of genes are involved in a single characteristic.



We inherit some of the genetic traits of our parents, and this is why we resemble them. In every cell of our body, we have 50% of each of our parent's genes*. In this way, all living things are survived by offspring which resemble them. In the early 20th century, scientists, while searching for materials containing genes, discovered that DNA is the material containing all genetic instructions.

DNA is a popular term in movies and TV dramas. The blockbuster film "Jurassic Park" in 1993 has become a household title with its intriguing story of bringing back the dinosaurs into a modern world, recreating them with DNA extracted from dinosaur blood found in a mosquito preserved inside some amber. In reality, it is impossible to revive dinosaurs with today's technology, but many people have come to learn that all the genetic information for any biological life form is stored in DNA.

DNA and the Bases

What exactly is this amazing DNA, then? DNA is an acronym for DeoxyriboNucleic Acid, and it consists of endlessly long, chain-like molecules with a sequence of "bases" on the line. There are four bases involved in DNA molecules: adenine (A), guanine (G), cytosine (C), and thymine (T). A sequence of these chemical bases represents genetic information. Changes in the sequence determine the characteristics of a living thing.

DNA consists of two strands, not one. The bases of either strand are determined by the bases of the other one: for example, if adenine (A) is on one strand, then thymine (T) is on the opposite strand; likewise, if guanine (G) is on one strand, cytosine (C) is on the other. Such a base pairing is called "complementary binding" of the bases. Since information is stored in two strands, immediate recovery of information after damage to the other strand is possible - a backup measure to protect vital genetic information.

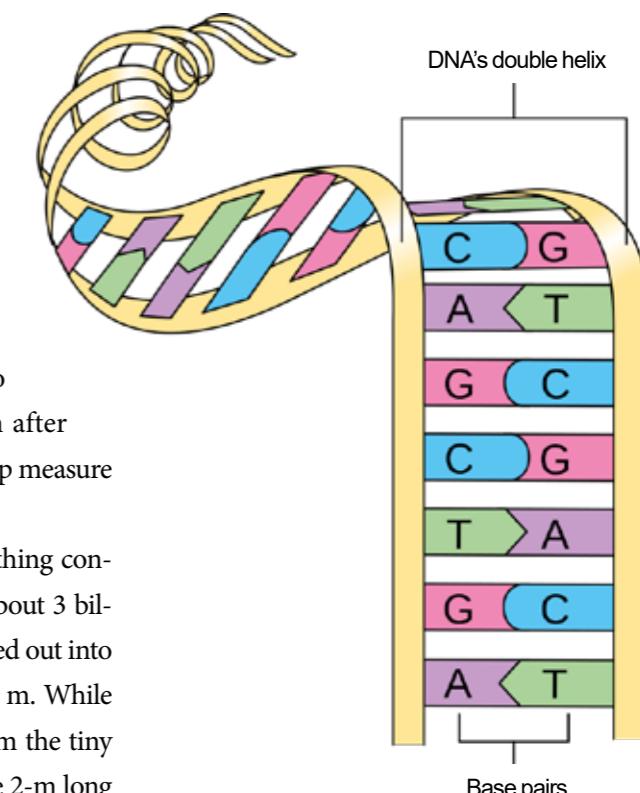
DNA that contains information on a living thing consists of a vast number of bases. For humans, about 3 billion bases are stored in one single cell. If stretched out into a single file, the line would easily span about 2 m. While this may not sound very long, it all comes from the tiny nucleus in a cell only 0.01 mm in diameter. The 2-m long DNA is folded up into a single cell, giving it an enormous amount of bio-information.

Isn't it amazing that such a vast amount of information about a living thing is stored in a combination of just four bases? But computers store information with a combination of just two digits: 0 and 1. Long before the invention of the computer, life might have been hinting at how we might store information in the future.

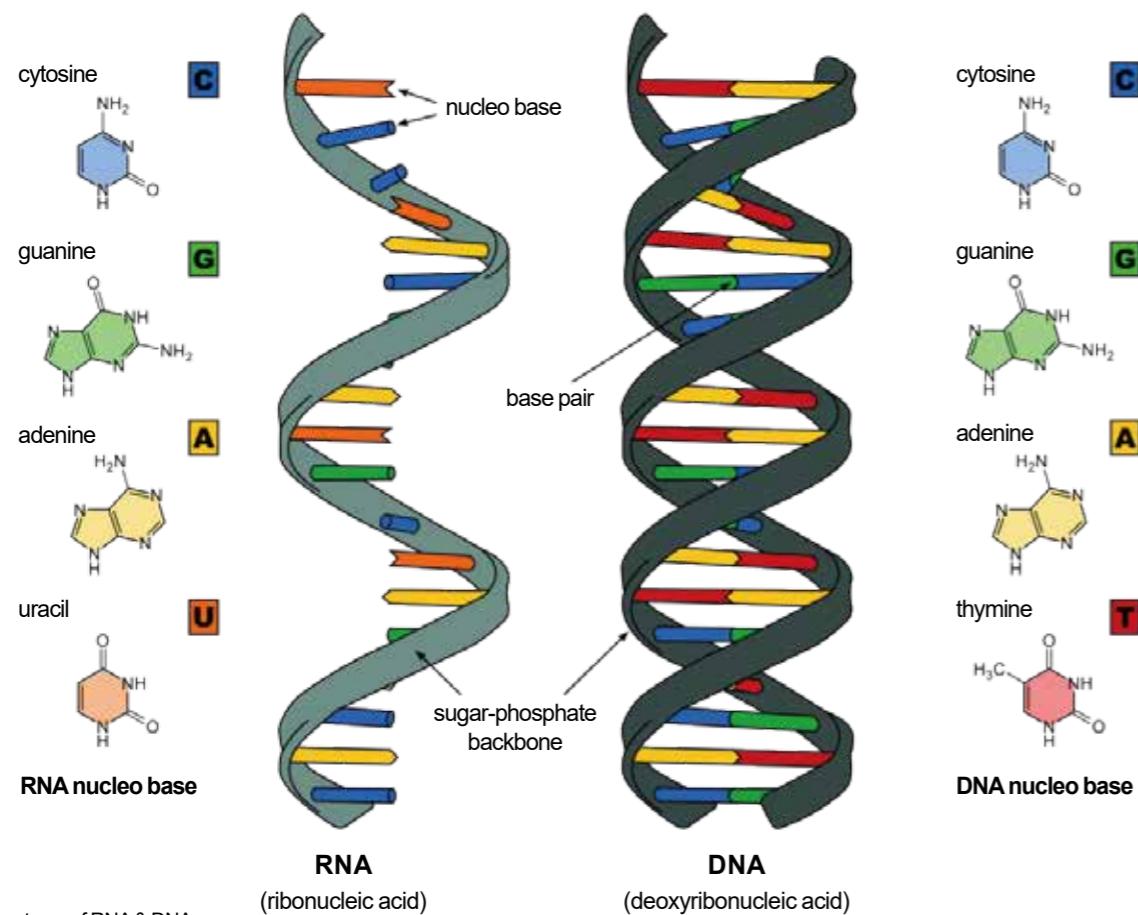
DNA and RNA

Another genetic material is known as RNA, which stands for Ribonucleic Acid. RNA is similar to DNA, but consists of slightly different molecules, and for bases, it has uracil (U) instead of thymine (T). Like thymine (T), uracil (U) pairs with adenine (A). Most importantly, DNA consists of two strands, and is thus very stable, whereas RNA is unstable since it has only one strand.

Instead, RNA plays multiple roles in the cell. One typical example: after duplicating some of the genetic information in the DNA, it



DNA has a double helix structure with two strands. Adenine (A) pairs with thymine (T); guanine (G) with cytosine (C) © Wikipedia



transfers necessary bits to produce protein - a process referred to as "transcription." In addition, it helps produce protein in the ribosome, an intracellular organelle that functions as a site for protein synthesis. It also delivers amino acids that are the chief components of proteins.

Recently, it has become known that RNA is far more versatile than we might think: it functions not just as a protein synthesizer, but also as an enzyme for specific reactions, or as an on/off switch for the functions of the genes.

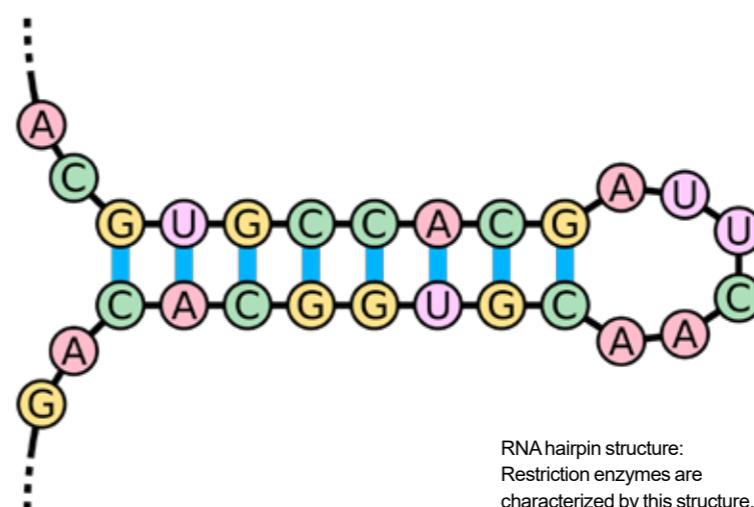
RNA Hairpin Structure

In the DNA structure, molecules that are supposed to be bound already are, leaving little room for other potentially interfering mol-

ecules. In contrast, RNA, having only one strand, has its bases exposed on one side. This means that other bases stick to the RNA, or sometimes its own bases bind together to form a loop. Imagine what happens with a roll of transparent box tape. It sticks to other stuff or sometime it sticks to itself and becomes entangled. This is similar to RNA.

With many bases that can bind to each other on its strand, RNA coils itself to form a loop and maintains its shape. A "palindrome" is a word or phrase that reads the same backward or forward, like the word 'level,' and if folded in the middle, two identical letters meet each other. RNA has a similar structure. The loop formed by the combinations of A-U's and G-C's resembles a hairpin: hence the term "hairpin structure".

RNA with a hairpin structure, as shown below, has two areas: an area where the bases bind to each other and the other where the bases do not, instead forming a loop. With this structure, RNA can perform unique functions. For example, it can work as a restriction enzyme that cuts up DNA fragments. When CRISPR, often referred to as gene scissors, are used to cut a specific sequence of bases, the hairpin structure plays a key role. For further details on CRISPR, see Part 3.



The changing definitions of "gene"

Initially, it was accepted that only one gene was involved in each genetic characteristic. With advancements in molecular biology, it is now known that the characteristics of a living thing are determined by the number and type of proteins that are synthesized.

The latest version of the definition of a gene is: "a sequence of DNA bases for the synthesis of one kind of protein," or "a sequence of DNA bases for control of the production of proteins."

Activity 1

Functions of RNA Bases

Human cells have chromosomes that contain genetic information, and scientists have discovered that DNA is the building block for chromosomes. The genetic information in DNA, which is copied into RNA to produce proteins, varies according to the sequence of the four RNA building bases of adenine (A), guanine (G), cytosine (C), and uracil (U). Each base has another base for pairing: adenine (A) binds with uracil (U), and guanine binds with cytosine (C). Thanks to this feature, RNA can store genetic information and generate specific reactions in the body of a living thing. Through the following role play, you will learn more about the roles of RNA bases.

What to prepare A4 paper, adhesive tape, pen

Activity

1 Pairing Game

① Get as many sheets of A4 paper as the number of participating students, and print A, G, C, and U, one on each paper, in big, bold letters. (For 20 participants, for example, have 5 sheets for each letter, making a total of 20 sheets)



② Stick the paper onto the front of the students' clothes.

③ Now the students are the RNA bases, moving around in the classroom freely. Base molecules also move around in the cell freely. Remember that A pairs with U, and C pairs with G.

④ With a "Go!" find a matching partner and hold hands. Each pair should hold left hands so that they face in the opposite direction to each other.



⑤ Repeat the game several times until the students become familiar with the fact that RNA bases pairs up only with their specific matching bases.

2

Make a Hairpin Structure

① Form a line in single file as below. Put your right hand on the right shoulder of the student in front. (The order shown here is based on a group of 20 participants. If there are fewer than 20, remove the letters at both ends accordingly.)

AGCGAAUGCUGACAUUCGCU



② Following the lead of the student at the head of the line, play choo-choo train slowly. This is the way the RNA strand moves freely in the cell. The lead student should move slowly to avoid breaking the train, which proceeds in a tight formation until a hairpin structure is formed.

③ The students' left hands are free. If you find a matching base in your vicinity, hold left hands.

- The binding force of the bases is weak. If only one or two pairs form, you should release your hand. Actually, RNA's weak bindings remain briefly and break in no time. Keep playing choo-choo train again.

- With at least three pairs the binding force becomes strong enough to maintain the pairing. The train then comes to a stop.



④ Check out the formation of the train. Make sure it forms a loop.

The loop-shaped RNA structure is referred to as a "hairpin structure." Thanks to this structure, RNA functions as an enzyme and helps CRISPR snip genes with precision.



⑤ Alter the sequence as below, and repeat a few more times

For Round 2

AAAAAGCCGUCGCGCUUUUG

For Round 3

GUUUUCCGGAGCCGAAAAU

PART 2

Recommended target
Middle school free semester

Relevant subjects
Science and human civilization,
3rd grade science

DNA Coding & Gene Recombination

As DNA contains all of an organism's genetic information, it is possible to alter the characteristics of a living thing by manipulating its DNA. For quite a long time, scientists have worked hard to engineer DNA, and have made significant achievements. One technique is referred to as "gene recombination," whereby a new gene is created by combining the whole or a part of the DNA taken from a specific cell with other DNA. This technique allows us to make pest-resistant, easy-to-grow corn, or microorganisms that produce essential substances to treat specific diseases.

The Coding in DNA

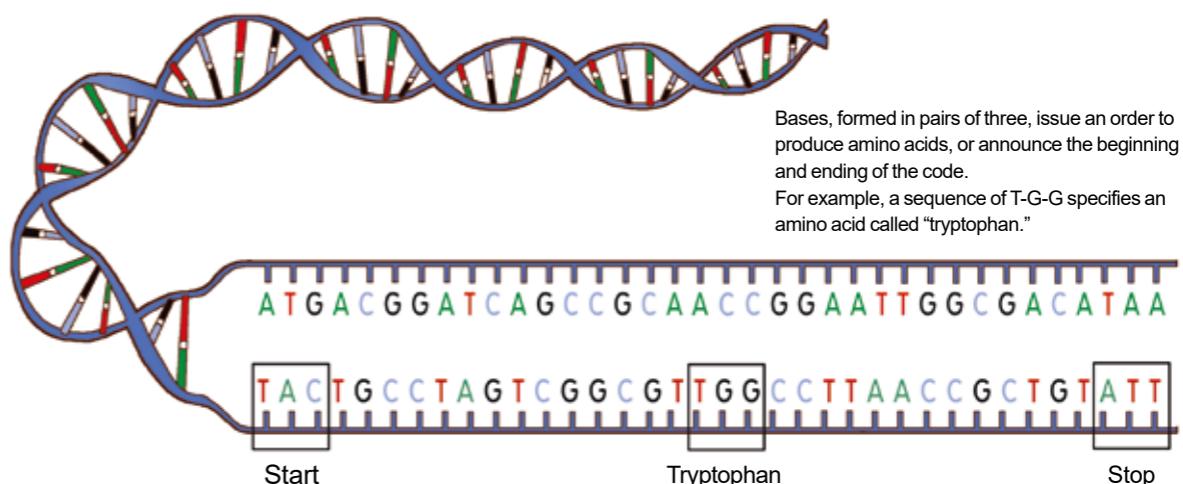
Thanks to scientists, we can now read DNA base sequences. At first, only a small number of DNA sequences could be analyzed, but gradually, it has become possible to analyze longer ones. The technique

for this analysis is highly complicated, so we will not discuss it here. Today, the technology has reached a point where if you put a DNA specimen into a certain device, the sequence will be analyzed automatically.

DNA sequence can be obtained in this way, but comes out as a kind of code with a mix of A, G, C, and T. Nothing can be obtained only with the sequence - you need to decipher the meaning. That's why sometimes the DNA sequence is referred to as a "code." Researchers have made every effort to break this code.

Of the many things that have been discovered, one is that a "codon" is a series of three adjacent bases, and each codon specifies one amino acid, which synthesize proteins. A codon is one of a total of 20 amino acids, and several amino acids link together, like a chain, to make protein. In short, with the code of codons, a variety of proteins can be synthesized.

However, not every codon can produce protein. In fact, only a fraction of the DNA makes protein. One codon, for example, demands, "Read off the code from here," and another commands, "End with the reading here." Some of them coordinate in determining the amount of protein to be produced.



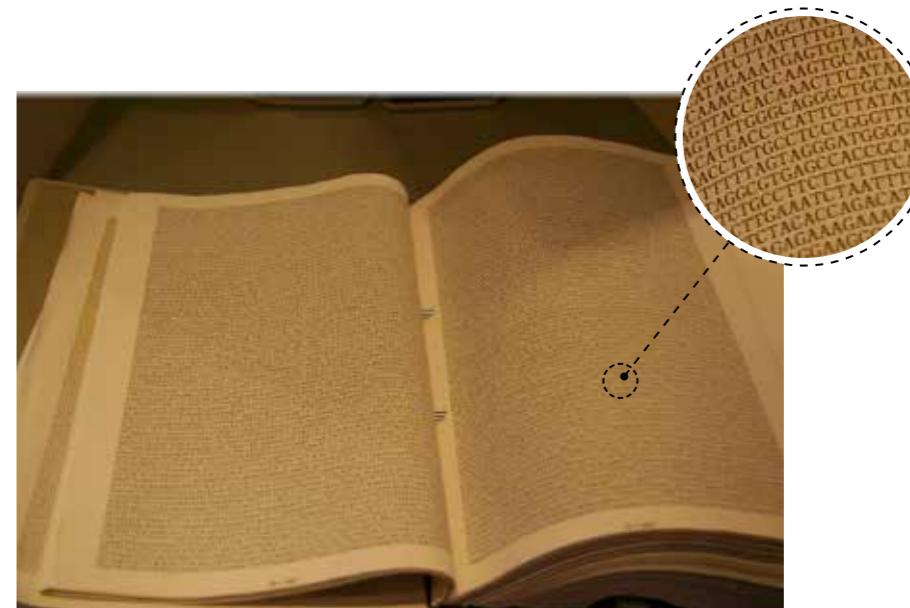
Things We can Do with Gene Recombination

True, there are so many things yet to be discovered, but we can still do so many things with what we already know. What can you do if you could manipulate genes whose specific functions you knew? Make a list of things you would like to do. Whether such a gene actually exists or not is not important - there are many more genes yet to be discovered

Genes That	Your Wishes
Decide hair color	
Build muscle	
Make you gain weight	

Remember: the best time to get the most out of genetic manipulation is when there is only one cell. For humans, this is when an egg inside the mother and sperm from the father unite to form a fertilized egg, while in the case of a plant, it's a seed. An adult human has approximately 100 trillion cells, and it is certainly impossible to correct the DNA in every cell one by one. So, if the genes are manipulated when there is only one cell, all the other cells created through cell division would carry the same genes.

For this reason, scientists have conducted many gene recombination experiments with single-cell organisms. For example, bacteria like *E. coli* are single-cell organisms. In addition, these germs undergo cell division every 20 minutes, resulting in an enormous number in



A book packed with the base sequences of human genome genes
©Ben Casey

just one day. Insert an insulin-producing gene into one *E. coli*, and, in a few days, you have a "factory" that produces insulin, an essential substance for treatment of diabetes.

Attempts to analyze all DNA in an organism are continuing. The first one completed was the bacteriophage X174 virus, which has only 5,386 base pairs, the smallest number of all organisms. Compare that to the approximately 5 million in a single *E. coli* cell, and the approximately 3 billion in a human cell.

The Human Genome Project was an effort to analyze all 3 billion pairs of human genetic information. With an investment of \$3 billion over a period of 10 years, it was the largest scientific undertaking since the moon landing. In the beginning, progress was slow, but this later picked up speed thanks to a newly introduced technique which eventually resulted in successful analysis of every base sequence in 2003.

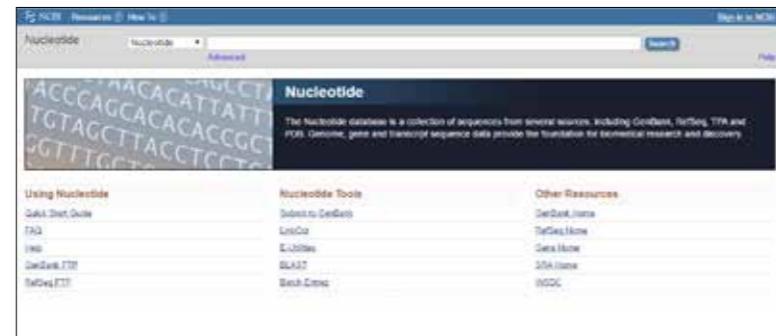
Information about DNA in today's world is overwhelming - more than enough. Therefore, rather than simply increasing the amount of information, now it is more important to hone existing skills towards identifying what is needed and how to use it.

Viewing the Nucleotide Sequence of a Specific Organism

If you would like to see the genetic coding of a specific organism, check out the website of the National Center for Biological Information (NCBI) for nucleotide sequences in DNA for a broad range of organisms. Follow these easy steps below:

1

Access the website at www.ncbi.nlm.nih.gov/nuccore.



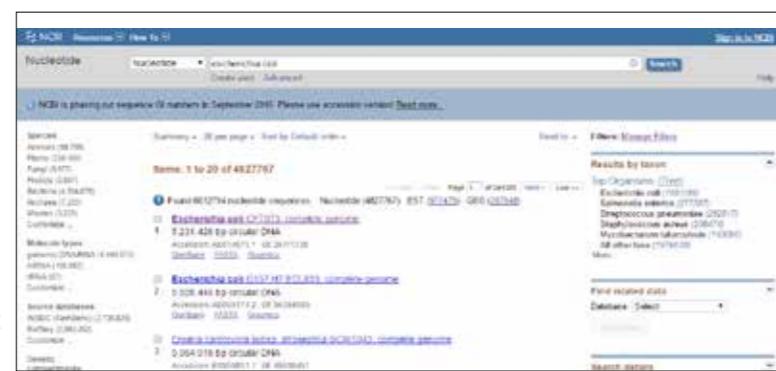
2

Type the scientific name of the organism in the search box and the DNA sequence information will display. For humans, type "homo sapiens," and "escherichia coli" for E. coli. Type "scientific name" for whatever organism you want to view in the search box to get the relevant name.

3

For example, let's check out the nucleotide sequence information for E. coli.

Type "escherichia coli" in the search box and click the "Search" button. A large number of E. coli strains may pop up on the screen. Simply click the first one at the top of the list, and ignore the rest.



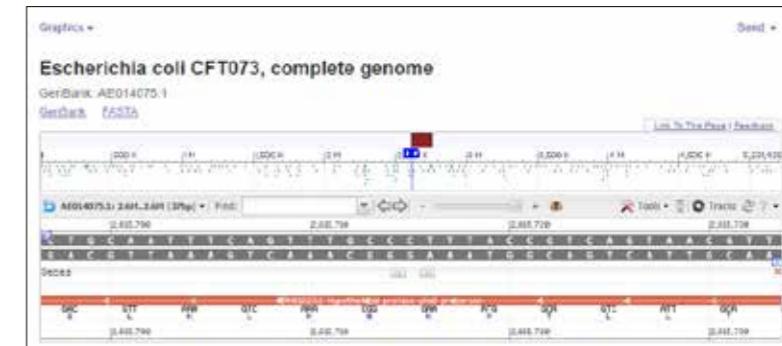
4

At the top left of the screen, you will see small links for "FASTA" and "Graphics." Clicking "Graphics" will display the screen shown below, which is a summarized view for large amounts of data. Move the - + control bar in the center to the far right for an enlarged view size of 100%.



5

Can you see the nucleotide sequence? This identifies the base sequence of E. coli. Go back to Step 4, and click "FASTA." Even with a fast computer, it may take quite a while until you get the view - after all, the number of genomes in the sequence of E. coli is well over 5 million.



FASTA ▾

Escherichia coli CFT073, complete genome

GenBank: AE014075.1

GenBank Graphics

```
>gi|26111730|gb|AE014075.1| Escherichia coli CFT073, complete genome
AGCTTTCACTCTGCTCAA...AGCTGTCTGA...TACAAATCTTAAACCAATAGCACTAAATACCTTAACCAA
TATAGCCATAGCGCA...AGCTGTCTGAGTA...AAATTAAATTCAGAGTAA...ACACATCCATGAA...CGCATTAACGA
CCACCATTACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCA
ATTACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCACTACCA
CCCGACACTGACAGTGCAGCTGGCGCTTTTTTGTGCAAGCAGAAAACCGCGACCTAGCTGGGGGTTCGGGAAG
CTTTCAGCTTGAGCGAGTTTAAACACCTTGTGATTGTTAAACACCTTGCAGCTGGGGGTTCGGGAACCTGCA
AAGTGTCAAAACAGAAATCAAAAGGGGGTCCAAATGGGGAGCGAAAAGAGCTTACGGCACACCGATGGAA
CTGTAATATACACATAGTATTTGGCGCAAATAACCGAAGCAGGTGTCTACAGAGAGAGGTAGACCA
ATAGGCTGTATTTGGAGAAAGCTGTGTGAGTGGGGAAAGTACCGATTCTGGAGCTGAATGCTGTGCA
ATCATATCCATATGCTGTGGAGATCGCGCAAAATGACCGTATCAGGCTTATGGGATACTGAAAGG
GAAAGCAGCTGTGATGCGCTTAAGAGCAGTTGGGTGATTGAAATTCAAATACAGGAACAGGAATCTG
TGGAGAGGGATTACGTGCAATACGGTGGGTAAAGAACACGGGAAGATACAGGAATCATAAAGCAACAGC
TTGAAGAGGATAAAATGGGGAGGAGCACTATCGATTCCCTACCGGGAGCGCGTTAACGGGCGTAAAGTA
AAGGAAGTGGATGCAAAATGTCAGATGTTGAGTGGGTGAGCTGGGAGCGCTGGTAAAGAGAGCCTTATAGG
```

Activity 2

Find a Specific Nucleotide Sequence!

Now you've learned how to search for the entire genetic sequence of an organism on the Internet. Isn't it amazing that all the information of a living thing is stored in DNA in the sequence of the four bases of A, G, T, and C? There are a huge number of genetic sequences in an organism. Locating a specific gene would take an enormous amount of time. When dealing with such vast amounts of data, a good computer program will help. Let's find a gene with specific functions in the bacteriophage phiX174 virus.

What to prepare PC with Internet connection

Activity

- 1 The characters on the next page are the entire nucleotide sequence of bacteriophage phiX174. This virus is one of the organisms that have the least number of the entire genetic sequence.
- 2 There is a series of “ATGGTTCG...” in the gene that represents “protein A” of bacteriophage phiX174. Find the matching character string on the next page.
- 3 Can you find it? It’s almost impossible with your eyes alone. But with the help of a computer’s search function, you can find it easily. Get a file from your teacher and open it with Windows Notepad. Press **Ctrl** + **F**, and the search box will open. Then type “ATGGTTCG” in the box, press “Find” and voila!
- 4 Have you found the character string? The string you found is the location where the virus makes protein A from the DNA information of bacteriophage phiX174. This virus produces a total of 11 proteins. Find the locations of other proteins listed on the next page.

Protein A	ATGAAATC...	Protein F	ATGTCTAA...
Protein B	ATGGAACA...	Protein G	ATGTTTCA...
Protein C	ATGAGAAA...	Protein H	ATGTTTGG...
Protein D	ATGAGTC...	Protein J	ATGTCTAA...
Protein E	ATGGTACG...	Protein K	ATGAGTCG...

5 Have you found all these genes? Probably not all of them - some are genes that cannot be found with only the given nucleotide sequence. For example, from your search, you get a total of two locations for protein D; three for protein F; and three for protein J. It is simply impossible to identify exact locations with only eight nucleotides in a sequence. Thus, the more of the sequence you know, the more accurately you can identify them. (Don't forget this! It is essential to understanding CRISPR.)

※ Find each of the following protein locations with the corresponding strings: Protein D with “ATGAGTCAA”; Protein F with “ATGTCTAATA”; Protein J with “ATGTCTAAAG”.

<The Entire Nucleotide Sequence of Bacteriophage phiX174>

※ For the genetic sequence of bacteriophage phiX174, visit this website: <http://www.ncbi.nlm.nih.gov/nuccore/216019?report=fasta>

PART 3

Recommended target
Middle school free semester

Relevant subjects
Science and human civilization,
3rd grade science

CRISPR: Genetic Scissors

It is known that genes can be found in DNA and RNA, and further studies have found that DNA and RNA sequences are the tools that store genetic information. Genetic sequences are as complex as cryptography, and many are still unknown. Yet, there are so many things that we can do with what is known to us now. If we know the gene of a substance that is rare in nature (a medicine, for example), we can mass-produce it by inserting the gene into *E. coli*.

Scissors That Cut Genes & A Glue That Pastes Them Together

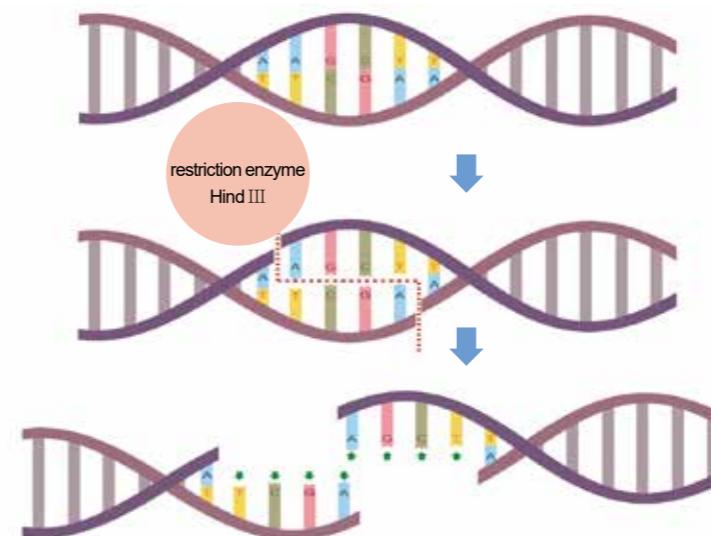
DNA or RNA are tiny molecules, hardly visible even under a microscope. The most dexterous surgeon won't be able to see them, let alone cut them up. In order to cut a molecule, we need "scissors" on the scale of molecular units. Continued scientific research has found that upon viral attack, bacteria protect themselves by producing specific enzymes* that cut out the foreign DNA. In other words, the bacterial self-defence mechanism is used as a kind of scissors to cut out genes. (When it comes to biology, there is no such thing as human creation. We only find things that already exist in nature and use them by altering as necessary.)

This enzyme does not cut up every sequence but restricts itself to specific ones: hence the name "restriction enzyme." For example, the restriction enzyme "EcoRI" found in *E. coli* cuts out the targeted DNA only when it encounters the sequence of "GAATTC." In general, a restriction enzyme recognizes six to eight genetic sequences, and there are several hundred restriction enzymes that have been found so far. In genetic recombination, the greater the number of restriction enzyme types, the greater the number of scissors available, and, in turn, the greater the number of tasks that can be accomplished.

* enzyme

A catalyst is a substance that increases the rate of a chemical reaction. With a catalyst, reactions occur explosively even in otherwise less likely chemical reactions. Proteins that act as a catalyst in living organisms are called enzymes.

A variety of chemical reactions (such as cutting, joining, and transferring) in organisms are caused by enzymes - like digestive enzymes that break down the food we eat to enable better nutritional absorption. Enzymes are involved in almost all reactions that occur in an organism.



The restriction enzyme, HindIII, when it encounters the genetic sequence of "AAGCTT," snips the DNA. ©Ben Casey

A joining enzyme known as "ligase" puts genes together. Restriction enzymes act like scissors, while ligases act like glue, so to speak. Unlike restriction enzymes, ligases do not recognize specific locations but simply join any loose DNA fragments together.

Therefore, precision in the use of the "scissors," rather than the "glue," is more important.

The Problem of Inaccurate Restriction Enzymes

Even with many restriction enzymes, you cannot cut the targeted DNA with 100% precision. In [Activity 2], with genetic sequences of only eight nucleotides in bacteriophage phiX174, which has the lowest number of nucleotide sequences of all organisms, multiple locations popped up as a result of your search. This means that using restriction enzymes as a DNA cutting tool may result in locations being cut despite the fact that they were not specifically targeted.

Earlier, we talked about the controversy surrounding GMOs. You may want to insert insect-resistant genes into the crops, but it is almost impossible to target the exact location of the genome. If inserted in the wrong locations, it is impossible to know all the potential side effects. There are many reasons for the controversy over GMOs, and inaccuracy in the technology of genetic recombination is one contributor.

+ Genetically Modified Organisms (GMOs)

A GMO is an organism whose genomes have been altered through genetic recombination techniques - mostly for food crops. South Korea imports GMO corn and soybeans to produce snack foods, beverages, and soy sauce. Since 2001, due to recurring controversy about the safety of GMOs, the government has implemented a Gene Recombinant Food Labeling System.

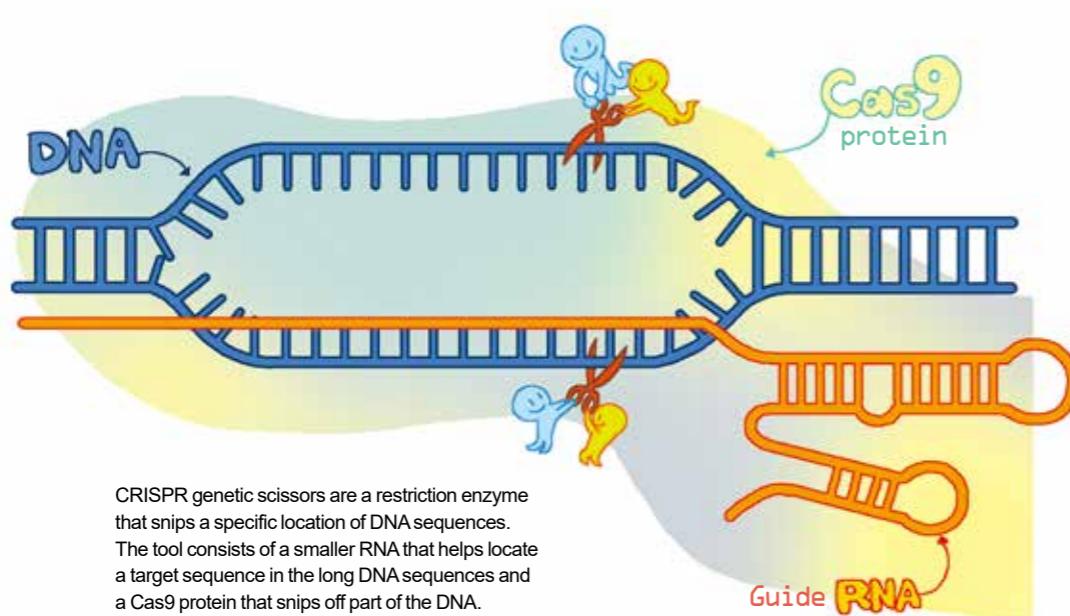
What if you had a precision tool such as genetic scissors that could be used to cut only the targeted locations? It would be a dream-come-true, and it has recently become available to biologists - the genetic scissors known as "CRISPR." The existence of these scissors was first discovered by scientists back in 1987, but only in 2012 were their potential uses found - the technology is less than 10 years old.

Protein + CRISPR = Genetic Scissors

CRISPR is an acronym for clustered regularly interspaced short palindromic repeats. In short, it is fragmented RNA with a certain palindrome structure.

Only when CRISPR joins with a protein called Cas9 does it function as genetic scissors. Shown below is the structure of CRISPR. Cas9 protein is a DNA-cutting enzyme, and a "guide" RNA helps locate a specific DNA. If a guide RNA sticks to the target location of DNA sequences, Cas9 protein cuts the DNA off.

With previous generation genetic scissors, there was always the possibility of mistakes as the scissors could recognize from four to twelve nucleotide sequences only. In contrast, as CRISPR recognizes 21 sequences, it can snip off any specific locations of the targeted DNA with precision.



It is virtually unlikely that there exist two or more sets of sequences that exactly match a string of 21 characters in the DNA sequences of an organism.

However, CRISPR is not a silver bullet as yet. In lab experiments, the tool often cuts in the wrong locations with unmatched sequences of the 21-character strings. Consequently, a procedure is essential for confirmation that the gene cutting has been accurate - another reason for co-development with other technologies.

The Future of Genetic Recombination Technology

All in all, thanks to CRISPR, the possibility of biological manipulation targeting an entire life form has opened up. It offers almost dream-like potential: of producing perfect GMOs, curing AIDS and cancer, and reviving extinct species.

As mentioned earlier, DNA sequences contain enormous amounts of data. In humans, there are 3 billion genome sequences in a single cell. It is impossible for people to analyze this sequence one by one - we simply need computers. As we have in our hands a perfect pair of genetic scissors like CRISPR, computer-aided collaboration between Big Data and the biosciences would be all the more important.

Yet, we still have a long way to go. Yes, it is now possible to snip off specifically-targeted genetic fragments, but inserting a specific gene into a specific location is another story - this is a time-consuming process and highly unlikely to succeed. Furthermore, we need to find more efficient ways to transfer extracellular CRISPR to intracellular locations.

There also remain ethical issues and side effects. As was the case when cloned animals appeared, genetic manipulation experiments on humans raise ethical concerns. Heated debates have already occurred, and progress in the biosciences should be made with consideration given to these ethical concerns.



May 2015 in New York, USA
GMO opposition citizens march.
© a katz/Shutterstock.com

Activity 3

Make Your Own CRISPR Game

To snip off a specific part of DNA, you only need to know approximately 20 nucleotide sequences before and after the target location. With this information, you can make a CRISPR that will cut the target area precisely. Use Scratch to create your own “CRISPR Game” where a CRISPR chases and snips a DNA on the run.

What to prepare PC or a laptop installed with Scratch (v. 2.0)

Activity

1 Mission

Create a project that meets the following conditions:

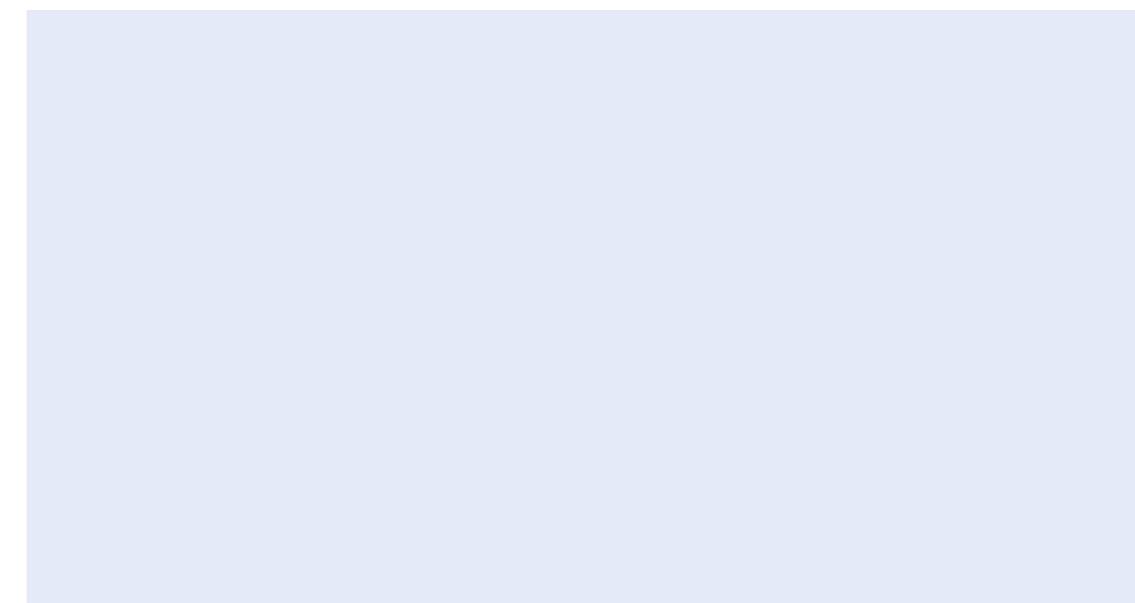
Conditions
① Use an event key and create a moving CRISPR
② Apply CRISPR's basic principles
③ Use your own personalized Sprite animation

2 Project Design

① Design your own unique CRISPR and DNA images.

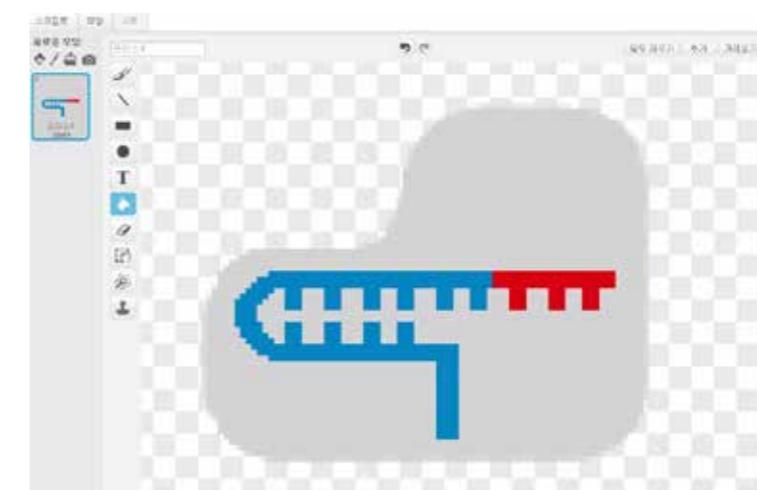


② Write an algorithm that makes CRISPR move upon mouse or keyboard command and reacts to an elusive DNA when encountered.



3 Programming

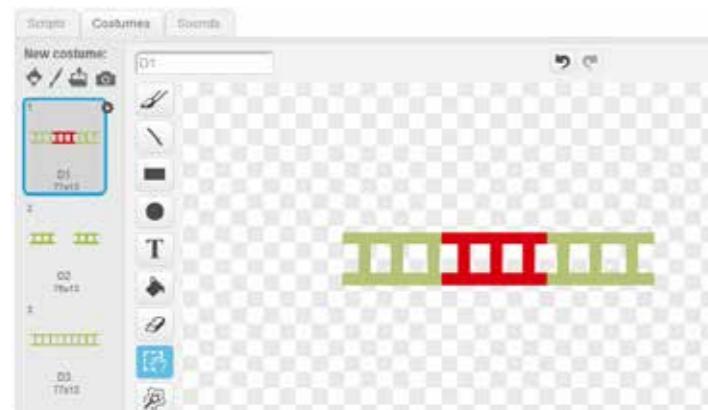
- ① To design a background wallpaper, select “Draw New Background”. Choose the color you want by clicking “Select Color,” and change it using “Coloring” from the menu on the left. Then, change the name to “Cell” in the background.
- ② To add a CRISPR Sprite, select “New Sprite Color” from “New Sprite,” and draw your own unique CRISPR using the brushes and shapes from the Paint.



③ Adjust the position and size of your Sprite in the background wallpaper with your mouse. Select “New Sprite Paint,” and add your own CRISPR Sprite using the drawing tools.



④ To add a DNA object, select “New Sprite Paint,” and add D1 Sprite using the drawing tools. Then add D2 Sprite to change the DNA into a snipped-off shape.

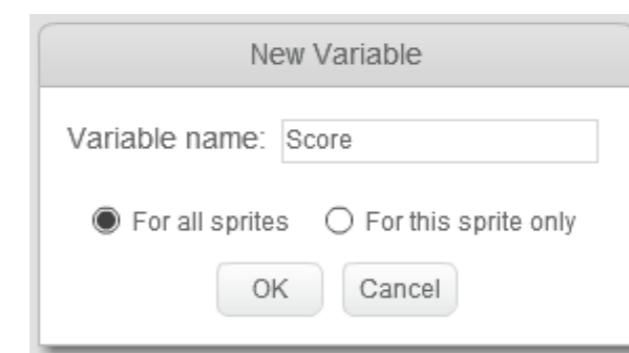


⑤ To add more fun to the game, make the CRISPR Sprite appear in unpredictable locations by setting x- and y-coordinate positions using random number blocks.

Set program for the CRISPR Sprite: “If the distance to the mouse pointer is greater than 5, then move in the direction of the pointer.” In this way, the mouse pointer does not necessarily have to be attached to the Sprite.



⑥ Create a variable for the score to raise it when CRISPR encounters the DNA Sprites. Click “Add New Variables” to add a variable to the data frame.

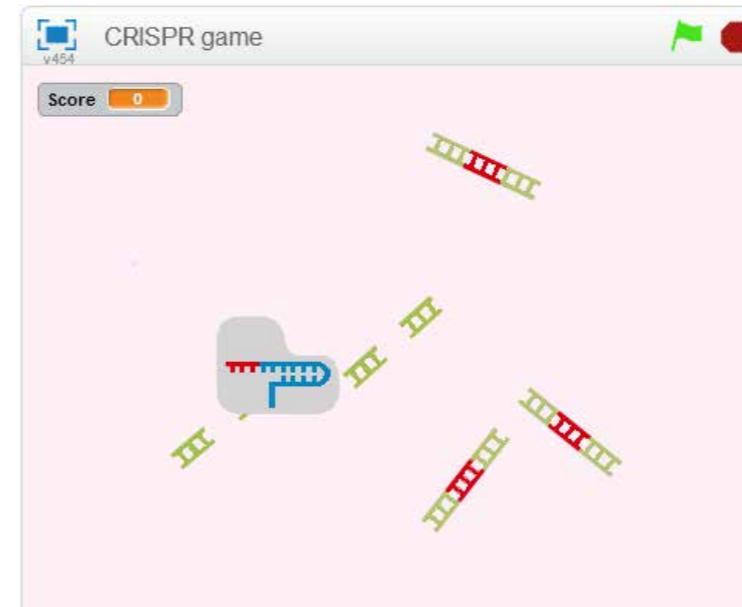


⑦ Reset the score to zero, and assemble the blocks into the shape of D1 Sprite.

Use the “random number” block to make D1 Sprite appear at an unpredictable location. Then, in order for it to move, assemble the “Move by 2” and “Bounce when touching the wall” blocks in the “Repeat infinitely” block.



⑧ Run the program to make sure there are no errors. Correct any errors that appear.

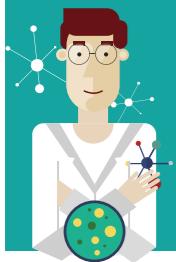


4 Select one of the following to change the program

- ① Create a program to end the game in 30 sec.
- ② Make the game more lively with sound effects.

5 Summarize the Activity

Describe briefly what you have learned from the activity.



Bioinformatics Specialists

A bioinformatics specialist studies how to organize, analyze and use biology-related data on computers. It's a typical convergent occupation, a combination of bioengineering and IT. Recently, as genetic engineering and other forms of cutting-edge biotechnology were designated as key national economic areas, venture companies related to the genetics industry have been emerging on a regular basis. In addition to the existing medicinal areas, other related sectors such as agriculture and the environment are expected to grow rapidly through application, and demand for bioinformatics specialists is anticipated to grow as well.



What do they do?

- In collaboration with specialists in medicine, pharmacy, and environmental engineering, bioinformatics specialists accumulate existing knowledge in the fields of genetic recombination and organism processes and use it to develop new bioengineering technology.
- Through DNA manipulation, they conduct experiments to explore new materials, or using somatic cells or the blood of lab animals, develop biomedicines, biochemicals, biofoods and other new products.
- By analyzing biological data on computers, they come up with new theories and methodologies.

Bioscience/ engineering	Bioscientist, bioengineer, new medicine developer, medical technology developer, stem cell researcher and developer, new biomaterial developer, bioengineering professor, medical department professor
Bioinformatics	Genetic information analyst, protein information analyst, genome researcher, interactome researcher

Occupations Related to CRISPR

Bioscience is a very important area to human health and longevity. Today, with the rush of ever-growing information flows from life-related research efforts such as the human genome project, the science of bioinformatics, which analyzes this data and its implications, has become increasingly important.



Related majors



It is necessary to have, among other things, basic knowledge about and understanding of the field of bioscience and its research activity. On top of this, you will need to be able to process bioscience-related data on a computer. Basic skills and knowledge related to databases, big data, and artificial intelligence are essential.



Required aptitude

- Knowledge of related sciences such as biology, engineering, medicine, and pharmacy is required, and clear understanding of the potential and limitations of the related areas is also needed.
- Continued research activity in the field of bioscience demands physical stamina, perseverance, and patience.
- Most bioengineers work as part of a team to carry out a project. Good social skills and the ability to cooperate with researchers in other areas are needed.
- In order to understand and apply natural laws and scientific research methods, you should have experience with problem solving, the ability to rationalize, and learn how to develop your observational skills.
- Math skills are essential for systematic analysis of life phenomena.



Expert interview

Thanks to the mysterious development process of organisms, humans are aware, develop patterns of behavior, personalities, and habits, and are able to take the actions necessary to survive. I was personally intrigued by this process, and here I am at the Laboratory of Genes and Development. Currently, I work on the role of genes and the environment in the development process, and how the interactions between the two are related to evolution. If you are someone who can endure boring experiment processes and rejoice in the results, you might have what it takes to be a future bioscientist.

Daehan Lee/PhD candidate at Laboratory of Genes & Development,
Dept. of Bioscience, Seoul National University; BSc from Dept. of Bioscience, SNU



CRISPR: Genetic Scissors



Written by
Chang-min Choi (Dongpae Middle School)

Product overview

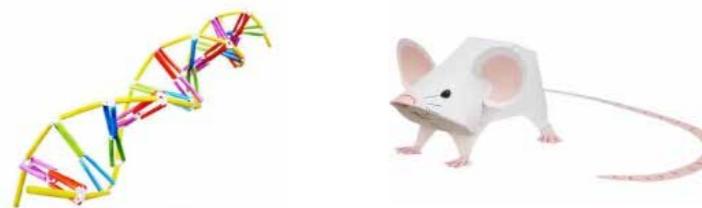
Product function

For elementary grade students, the functions and applications of CRISPR can be illustrated with a product that mimics the action of cutting DNA with CRISPR, and genetic recombination.

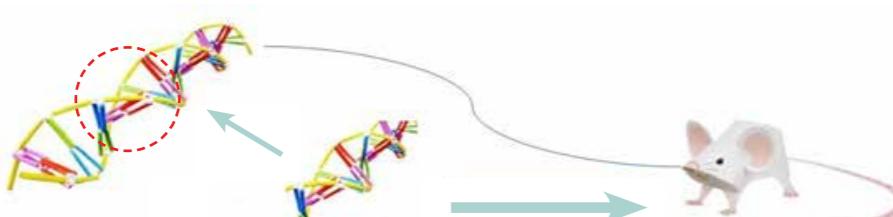
How it works

RGB (red, green, blue) LEDs are connected to the model double helix DNA structure, and inserted into the model mouse. The color of the LEDs changes with changes in the electrical resistance when other double helix DNA structures are connected.

Product picture



Product structure



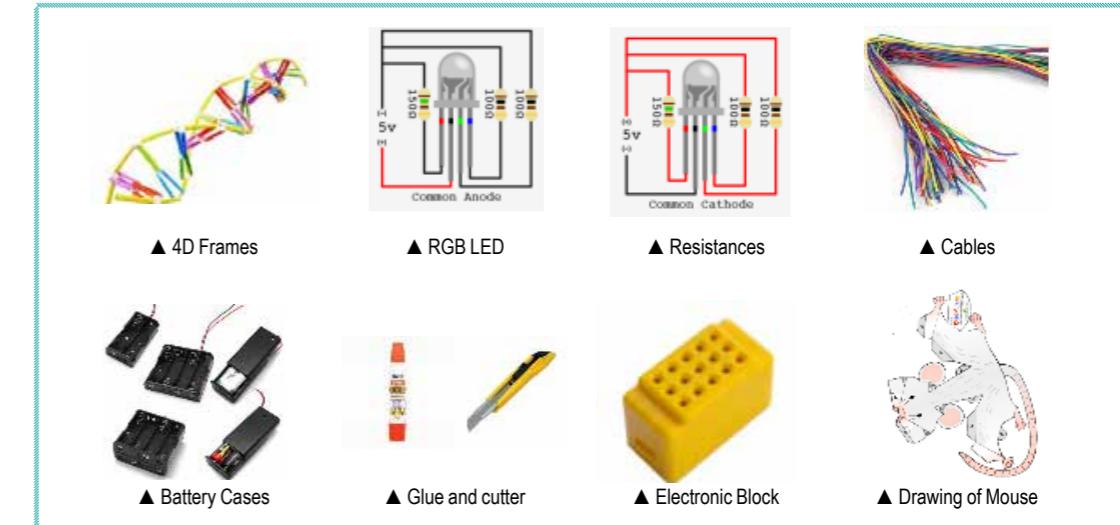
- Part of the DNA models changed
- The replaced DNA is assumed to be the gene that determines the color of the mouse

- LED color changes when part of the DNA model is replaced



Production overview

- Production time : about 2 hours
- Needed materials and tools



Key production principles

Genetic Recombination, LED Lighting through Resistance Change

Use 4D frames and RGB LED to confirm phenotypic changes from the CRISPR genetic recombination. Confirm the changes in LED color due to voltage drop*. Understand the CRISPR genetic recombination by observing these changes.

* Voltage Drop: a phenomenon that occurs when the voltage of a current decreases

What requires attention in production

- To confirm the LED color change from the voltage drop induced by resistance, connect the LED pins to the electronic block correctly.
- Cut out the model mouse drawing with care.
- When making the gene model that is to be recombined using 4D frames, the order must be in colors that are different from the DNA model being recombined.

Required knowledge and function

- Double helix DNA structure
- Gene cutting and recombination using CRISPR
- Voltage drop through resistance change
- Safe tool handling

SW Education Module Textbook

① Artificial Intelligence

② Driverless Vehicles

① IoT(Internet of Things)

④ Virtual Reality

⑤ CRISPR

⑥ Space Launch Vehicles

⑦ Natural Disasters

⑧ Smart Medicine

⑨ Game Engines

⑩ Sports Statistics

Problem-Solving Activities for Computational Thinkers ⑤ CRISPR

Third Generation Genetic Scissors: CRISPR

Issued on February 23, 2017

Issued by Ministry of Science, ICT and Future Planning, Korea Foundation for the Advancement of Science and Creativity

Written by Jeong Hoon Kim

Planned by Ji Young Park, Jong Keol Son, Hye Ja Kim (MSIP)

Yun Jeong Kim, Se Man Park, Oksu Hong, Hannah Jung (KOFAC)

Reviewed by Hyeon Ja Jeon, Seok Jin Lee, Hyeon Cheol Kim

Published by Doo Hui Kim

Publisher Dongascience Inc.

Responsible editor Kyeong Ae Chang

Inquiries Office of Software Education, Korea Foundation for the Advancement of Science and Creativity

Email swedu@kofac.re.kr

Phone (02)559-3841~9

Address (135-867) 602 Seolleung-ro, Gangnam-gu, Seoul (Samreung Building, Samseong-dong)

Homepage www.kofac.re.kr