

CRISPR/Cas9

Ma Hongbao¹, Margaret Young²

¹ Brooklyn, NY 11212, USA; ² Cambridge, MA 02138, USA
ma8080@gmail.com

Abstract: Clustered regularly interspaced short palindromic repeats (CRISPR) are certain prokaryotic DNA segments that contain short repetitions of DNA sequences. Each repetition is followed by some short segments related to a bacteriophage virus or plasmid. The CRISPR/Cas system is a prokaryotic immune system. CRISPR spacers recognize and cut the exogenous DNA for an immunological function. CRISPRs exist in the bacterial genomes. The Cas9 nuclease and guide RNAs can cut genome in certain location and remove the existing genes. CRISPRs can be used for genome editing and gene regulation. Using CRISPR interference technique can alter the germline of animals and plants.

[Ma Hongbao, Margaret Young. **CRISPR/Cas9**. *Researcher* 2017;9(8):71-80]. ISSN 1553-9865 (print); ISSN 2163-8950 (online). <http://www.sciencepub.net/researcher>. 11. doi:[10.7537/marsrj090817.11](https://doi.org/10.7537/marsrj090817.11).

Key words: clustered regularly interspaced short palindromic repeats (CRISPR); Cas9, gene editing, DNA; bacteriophage; immune system; life

The genome is the genetic material of an organism, which consists of DNA (or RNA in RNA viruses). The genome includes both the genes (the coding regions) and the noncoding DNA. Genome editing is a technique in which DNA is deleted, inserted or replaced in the genome of an organism using engineered nucleases that create site-specific double-strand breaks at desired locations in the genome. The induced double-strand breaks are repaired through nonhomologous end-joining (NHEJ) or homologous recombination (HR) which resulting in targeted mutations. NHEJ uses a variety of enzymes to directly join the DNA ends in a double-strand break and in HDR a homologous sequence is utilized as a template for regeneration of missing DNA sequence at the break point. Clustered regularly interspaced short palindromic repeats (CRISPR) are certain prokaryotic DNA segments that contain short repetitions of DNA sequences. Each repetition is followed by some short segments related to a bacteriophage virus or plasmid. The CRISPR/Cas system is a prokaryotic immune system. CRISPR spacers recognize and cut the exogenous DNA for an immunological function. CRISPRs exist in the bacterial genomes. The Cas9 nuclease and guide RNAs can cut genome in certain location and remove the existing genes. CRISPRs can be used for genome editing and gene regulation. Using CRISPR interference technique can alter the germline of animals (Ma, et al, 2016).

A gene knockout is a genetic technique in which one of an organism's genes is made inoperative. Also known as knockout organisms or simply knockouts, they are used in learning about a gene that has been sequenced, but which has an unknown or incompletely known function. Researchers draw inferences from the difference between the knockout organism and normal

individuals. The term also refers to the process of creating such an organism, as in "knocking out" a gene. The technique is essentially the opposite of a gene knockin. Knocking out two genes simultaneously in an organism is known as a double knockout. Similarly the terms triple knockout and quadruple knockouts are used to describe three or four knocked out genes, respectively. This article introduces recent research reports as references in the related studies (Ma et al, 2015).

Clustered regularly interspaced short palindromic repeats (CRISPR) are certain prokaryotic DNA segments that contain short repetitions of DNA sequences and each repetition is followed by short segments of spacer DNA. Each repetition is followed by some short segments related to a bacteriophage virus or plasmid. CRISPRs are found in approximately 40% of sequenced bacterial genomes and 90% of sequenced archaea. The CRISPR/Cas system is a prokaryotic immune system that digests the genes of invading viruses or plasmids.

Naturally the CRISPR spacers in bacteria recognize and cut the exogenous DNA for an immunological reaction. CRISPRs exist in the bacterial genomes. The Cas9 nuclease guided by guide RNAs (gRNA) can cut genome in certain location and remove the existing genes that match the gRNA. Using CRISPR interference technique can change the germline of animals and plants.

Naturally the Cas9 is an RNA-guided DNA endonuclease in bacteria. Cas9 can find and cleave foreign DNA, such as invading bacteriophage DNAs or plasmid DNAs. If the DNA substrate is complementary to the gRNA, Cas9 cleaves the DNA molecule.

CRISPR/Cas9 is a very useful genome editing

tool. The Cas9 endonuclease is a four-component system including two small RNA molecules and a DNA enzyme. By manipulating the nucleotide sequence of the guide RNA, the artificial Cas9 system could be programmed to target any sequence in DNA for cleavage. This technological advance has fueled efforts to edit genomes with the re-engineered CRISPR/Cas9 system. CRISPR has been modified to make programmable transcription factors that allow scientists to target and activate or silence specific genes. Libraries of tens of thousands of guide RNAs are available.

CRISPR's function is related to the cas genes. The cas genes express a helicase that unwinds DNA and a nuclease that digests DNA. CRISPR sequences directed cas enzymes to degrade viral DNA. Bacteria use spacers in their immune defenses. In 2015, the nuclease Cpf1 was discovered in the CRISPR/Cpf1 system of the bacterium *Francisella novicida*.

CRISPR repeats about 24 to 48 bp that show some dyad symmetry normally in a secondary structure as a hairpin. Repeats are separated by spacers. Some CRISPR spacer sequences exactly match sequences from plasmids and phages and some other spacers match the prokaryote's genome as self-targeting spacers. New spacers can be added rapidly as part of the immune response to phage infection. Small clusters of cas genes are often located next to CRISPR repeat-spacer arrays.

The 4 stages of CRISPR immunity: (1) Invading DNA is recognized and cleaved to protospacer by Cas1 and Cas2; (2) Protospacer is ligated to the direct repeat adjacent to the leader sequence; (3) Single strand extension repairs the CRISPR and duplicates the direct repeat; (4) The primary CRISPR transcript is cleaved by cas genes to produce crRNAs.

When a microbe is invaded by a virus, the first stage of the immune response is to capture viral DNA and insert it into a CRISPR locus in the form of a spacer.

Cas1 proteins have diverse amino acid sequences, which are metal-dependent nucleases/integrases that bind to DNA in a sequence-independent manner. Cas2 proteins have been characterized and possess either ssRNA- or dsDNA- specific endoribonuclease activity.

The PAM sequence appears to be important during spacer insertion in type I-E systems. That sequence contains a strongly conserved final nucleotide (nt) adjacent to the first nt of the protospacer. This nt becomes the final base in the first direct repeat. This suggests that the spacer acquisition machinery generates single-stranded overhangs in the second-to-last position of the direct repeat and in the PAM during spacer insertion. However, not all CRISPR-Cas systems appear to share this mechanism as PAMs in other organisms do not show the same

level of conservation in the final position. It is likely that in those systems, a blunt end is generated at the very end of the direct repeat and the protospacer during acquisition. Multiple CRISPRs contain many spacers to the same phage.

CRISPR-RNA (crRNA) guides the Cas nuclease to the target during the interference step. The crRNA is initially transcribed as part of a single long transcript encompassing much of the CRISPR array. This transcript is then cleaved by Cas proteins to form crRNAs. CrRNAs associate with Cas proteins to form ribonucleotide complexes that recognize foreign nucleic acids. CrRNAs show no preference between the coding and non-coding strands, which is indicative of an RNA-guided DNA-targeting system. All crRNAs contain a spacer sequence and some portion of the repeat at one or both ends. It is the partial repeat sequence that prevents the CRISPR-Cas system from targeting the chromosome as base pairing beyond the spacer sequence signals self and prevents DNA cleavage. RNA-guided CRISPR enzymes are classified as type V restriction enzymes.

CRISPRs are widely distributed among bacteria and archaea and show some sequence similarities, and they are evolutionarily conserved and cluster into related types. Many CRISPRs show signs of a conserved secondary structure. The basic model of CRISPR evolution is newly incorporated spacers driving phages to mutate their genomes to avoid the bacterial immune response, creating diversity in both the phage and host populations. Through the CRISPR/Cas mechanism, bacteria can acquire immunity to certain phages and thus halt further transmission of targeted phages. CRISPR/Cas-based RNA-guided nucleases can be used to target virulence factors, genes encoding antibiotic resistance, and other medically relevant sequences of interest. CRISPR sequences directed Cas enzymes to degrade viral DNA.

The CRISPR/Cas9 system consists of two key molecules that introduce a change into the DNA: (1) The enzyme Cas9, which acts as a pair of molecular scissors that can cut the two strands of DNA at a specific location in the genome so that bits of DNA can then be added or removed. (2) The guide RNA (gRNA), which consists of a small piece of pre-designed RNA sequence (about 20 bases long) located within a longer RNA scaffold. The scaffold part binds to DNA and the pre-designed sequence guides Cas9 to the right part of the genome. This makes sure that the Cas9 enzyme cuts at the right point in the genome.

The guide RNA is designed to find and bind to a specific sequence in the DNA. The guide RNA has RNA bases that are complementary to those of the target DNA sequence in the genome. The guide RNA will only bind to the target sequence and no other regions of the genome. The Cas9 follows the guide

RNA to the same location in the DNA sequence and makes a cut across both strands of the DNA. Spacer DNA is a region of non-coding DNA between genes. In bacteria, spacer DNA sequences are only a few nucleotides long, and in eukaryotes they can be extensive and include repetitive DNA, comprising the majority of the DNA of the genome. In ribosomal DNA the spacers are within or between gene clusters that are called internal transcribed spacer and external transcribed spacers, respectively. In animal the mitochondrial DNA genes generally have very short spacers. In fungi mitochondrial DNA spacers are common and variable in length, and they may also be mobile.

Repeated sequences are DNA or RNA multiple copies throughout the genome. In many organisms, a significant fraction of the genomic DNA is highly repetitive, with over two-thirds of the sequence consisting of repetitive elements in human. The repetitive DNA segments could be remainders from past evolution or autonomous self-replicating sequences hacking the cell machinery to proliferate. The disposition of repetitive elements consists either in arrays of tandemly repeated sequences, or in repeats dispersed throughout the genome. In human, several classes of repetitive elements present a high tendency for co-localization within the nuclear space, suggesting that DNA repeats positions can be used by the cell as a genome folding map.

The immune system is a host defense system comprising many biological structures and processes within an organism that protects against disease. To function properly, an immune system must detect a wide variety of agents, known as pathogens, from viruses to parasitic worms, and distinguish them from the organism's own healthy tissue. In many species, the immune system can be classified into subsystems, such as the innate immune system versus the adaptive immune system, or humoral immunity versus cell-mediated immunity. In humans, the blood-brain barrier, blood-cerebrospinal fluid barrier, and similar fluid-brain barriers separate the peripheral immune system from the neuroimmune system which protects the brain.

In 2015, the nuclease Cpf1 was discovered in the CRISPR/Cpf1 system of the bacterium *Francisella novicida*. Cpf1 showed several key differences to Cas9 including, such as: (1) causing a staggered cut in double stranded DNA as opposed to the blunt cut produced by Cas9, (2) relying on a T rich Protospacer adjacent motif, (3) requiring only a CRISPR RNA (crRNA) for successful targeting.

Clustered Regularly Interspaced Short Palindromic Repeats from *Prevotella* and *Francisella* 1 (CRISPR/Cpf1) is a DNA-editing technology analogous to the CRISPR/Cas9 system. Cpf1 is an

RNA-guided endonuclease of a class II CRISPR/Cas system found in *Prevotella* and *Francisella* bacteria, which prevents genetic damage from viruses. Cpf1 genes are associated with the CRISPR locus coding for an endonuclease that use a guide RNA to find and cleave viral DNA. The Cpf1-crRNA complex cleaves target DNA or RNA by identification of a protospacer adjacent motif 5'-YTN-3' (where Y is a pyrimidine and N is any nucleobase) or 5'-TTN-3', in contrast to the G-rich PAM targeted by Cas9. After identification of PAM, Cpf1 introduces a sticky-end-like DNA double-stranded break of 4 or 5 nucleotides overhang. Cpf1 leaves one strand longer than the other, creating sticky ends and Cas9 cuts both strands in a DNA molecule at the same position, leaving behind blunt ends. Using Cpf1 it is more able to insert new sequences at the cut site.

CRISPR/Cpf1 systems activity has three stages: (1) Adaptation: Cas1 and Cas2 proteins facilitate the adaptation of small fragments of DNA into the CRISPR array. (2) Formation of crRNAs: processing of pre-cr-RNAs producing of mature crRNAs to guide the Cas protein. (3) Interference: the Cpf1 is bound to a crRNA to form a binary complex to identify and cleave a target DNA sequence.

CRISPR repeats range in size from 24 to 48 base pairs. Small clusters of *cas* genes are often located next to CRISPR repeat-spacer arrays. Comparative genomics identified multiple *cas* genes have as many as 45 *cas* gene families. The classification groups CRISPR-Cas systems are classified as 2 classes. Class 1 systems use a complex of multiple Cas proteins to degrade foreign nucleic acids. Class 2 systems use a single large Cas protein for the same purpose. Class 1 is divided into types I, III, and IV; class 2 is divided into types II and V. The five system types are divided into 16 subtypes. Each type and most subtypes are characterized by a signature gene found exclusively in the category. Classification is also based on the complement of *cas* genes that are present. Most CRISPR-Cas systems have a Cas1 protein. The phylogeny of Cas1 proteins generally agrees with the classification system. Many organisms contain multiple CRISPR-Cas systems suggesting that they are compatible and may share components. The sporadic distribution of the CRISPR/Cas subtypes suggests that the CRISPR/Cas system is subject to horizontal gene transfer during microbial evolution.

In 2005, three independent research groups showed that some CRISPR spacers are derived from phage DNA and extrachromosomal DNA such as plasmids. In effect, the spacers are fragments of DNA gathered from viruses that previously tried to attack the cell. The source of the spacers is a sign that the CRISPR/*cas* system could have a role in adaptive immunity in bacteria. All the three studies proposing

this idea were initially rejected by so called leading journals but were eventually published in other journals. This also showed that the so called leading journals do not always know what works are important and many important achievements were published in so called normal journals.

In the CRISPR-Cas system the spacers produce small RNA guides to target RNA transcribed from viral DNA, analogous to the RNA interference system used by eukaryotic cells. Others hypothesized that CRISPR sequences directed Cas enzymes to degrade viral DNA.

Jennifer Doudna and Emmanuelle Charpentier studied a simpler CRISPR system from *Streptococcus pyogenes* that relies on the protein Cas9. The Cas9 endonuclease is a four-component system that includes two small RNA molecules. In 2012, they re-engineered Cas9 endonuclease into a more manageable two-component system by fusing the two RNA molecules into a single-guide RNA that, when mixed with Cas9, could find and cut the DNA target

specified by the guide RNA. By manipulating the nucleotide sequence of the guide RNA, the artificial Cas9 system could be programmed to target any sequence in DNA for cleavage. This technological advance has fueled efforts to edit genomes with the re-engineered CRISPR/Cas9 system.

Feng Zhang and George Church simultaneously described genome editing in human cell cultures using CRISPR/Cas9 systems for the first time. It has since been used in a wide range of organisms, including baker's yeast (*Saccharomyces cerevisiae*), zebrafish (*D. rerio*), fruit flies (*Drosophila melanogaster*), axolotl (*A. mexicanum*), nematodes (*C. elegans*), mice, monkeys, human embryos and plants, etc. CRISPR has been modified to make programmable transcription factors that allow us to target and activate or silence specific genes. Libraries of tens of thousands of guide RNAs are available. .

Table 1 shows the signature genes and their putative functions for the major and minor CRISPR-cas types.

Table 1. Signature genes and their putative functions for the major and minor CRISPR-cas types

Class	Cas type	Signature protein	Function
1	I	Cas3	Single-stranded DNA nuclease (HD domain) and ATP-dependent helicase
	IA	Cas8a, Cas5	Subunit of the interference module. Important in targeting of invading DNA by recognizing the PAM sequence
	IB	Cas8b	
	IC	Cas8c	
	ID	Cas10d	contains a domain homologous to the palm domain of nucleic acid polymerases and nucleotide cyclases
	IE	Cse1, Cse2	
	IF	Csy1, Csy2, Csy3	Not determined
	IU	GSU0054	
	III	Cas10	Homolog of Cas10d and Cse1
	IIIA	Csm2	Not Determined
IIIB	Cmr5	Not Determined	
2	IIIC	Cas10 or Csx11	
	IIID	Csx10	
	IV	Csfl	
	IVA		
	IVB		
	II	Cas9	Nucleases RuvC and HNH together produce DSBs, and separately can produce single-strand breaks. Ensures the acquisition of functional spacers during adaptation.
2	IIA	Csn2	Ring-shaped DNA-binding protein. Involved in primed adaptation in Type II CRISPR system.
	IIB	Cas4	Not Determined
	IIC		Characterized by the absence of either Csn2 or Cas4
	V	Cpf1, C2c1, C2c3	Nuclease RuvC. Lacks HNH.
	VI	C2c2	

A gene signature is a group of genes in a cell whose combined expression pattern is uniquely characteristic of a biological phenotype condition. The phenotypes that may theoretically be defined by a gene expression signature range from those that are used to differentiate between different subtypes that predict the survival or prognosis of an individual characteristic.

The CRISPR technology had been used to functionally inactivate genes in human cell lines and cells, to study *Candida albicans*, to modify yeasts used to make biofuels and to genetically modify crop strains, etc. CRISPR can also be used to change mosquitos so

they cannot transmit diseases such as malaria. CRISPR-based re-evaluations of claims for gene-disease relationships have led to the discovery of potentially important anomalies. CRISPR/Cas9 genome editing is carried out with a Type II CRISPR system. When utilized for genome editing, this system includes Cas9, crRNA, tracrRNA along with an optional section of DNA repair template that is utilized in either Non-Homologous End Joining (NHEJ) or Homology Directed Repair (HDR).

Table 2 shows the major components of CRISPR/Cas9.

Table 2. Major components of CRISPR/Cas9

Component	Function
crRNA	Contains the guide RNA that locates the correct section of host DNA along with a region that binds to tracrRNA (generally in a hairpin loop form) forming an active complex.
tracrRNA	Binds to crRNA and forms an active complex.
sgRNA	Single guide RNAs are a combined RNA consisting of a tracrRNA and at least one crRNA
Cas9	Protein whose active form is able to modify DNA. Many variants exist with differing functions (i.e. single strand nicking, double strand break, DNA binding) due to Cas9's DNA site recognition function.
Repair template	DNA that guides the cellular repair process allowing insertion of a specific DNA sequence

CRISPR/Cas9 often employs a plasmid to transfect the target cells. The main components of this plasmid are displayed in the image and listed in the table. The crRNA needs to be designed for each application as this is the sequence that Cas9 uses to identify and directly bind to the cell's DNA. The crRNA must bind only where editing is desired. The repair template is designed for each application, as it must overlap with the sequences on either side of the cut and code for the insertion sequence. Multiple crRNAs and the tracrRNA can be packaged together to form a single-guide RNA (sgRNA). This sgRNA can be joined together with the Cas9 gene and made into a plasmid in order to be transfected into cells.

CRISPR/Cas9 offers a high degree of fidelity and relatively simple construction. It depends on two factors for its specificity: the target sequence and the PAM. The target sequence is 20 bases long as part of each CRISPR locus in the crRNA array. A typical crRNA array has multiple unique target sequences. Cas9 proteins select the correct location on the host's genome by utilizing the sequence to bond with base pairs on the host DNA. The sequence is not part of the Cas9 protein and as a result is customizable and can be independently synthesized.

The PAM sequence on the host genome is recognized by Cas9. Cas9 cannot be easily modified to recognize a different PAM sequence. However this is not too limiting as it is a short sequence and nonspecific (e.g. the SpCas9 PAM sequence is 5'-

NGG-3' and in the human genome occurs roughly every 8 to 12 base pairs). Once these have been assembled into a plasmid and transfected into cells the Cas9 protein with the help of the crRNA finds the correct sequence in the host cell's DNA and – depending on the Cas9 variant – creates a single or double strand break in the DNA.

Properly spaced single strand breaks in the host DNA can trigger homology directed repair, which is less error prone than the non-homologous end joining that typically follows a double strand break. Providing a DNA repair template allows for the insertion of a specific DNA sequence at an exact location within the genome. The repair template should extend 40 to 90 base pairs beyond the Cas9 induced DNA break. The goal is for the cell's HDR process to utilize the provided repair template and thereby incorporate the new sequence into the genome. Once incorporated, this new sequence is now part of the cell's genetic material and passes into its daughter cells.

It can use viral or non-viral systems for delivery of the Cas9 and sgRNA into target cells. Electroporation of DNA, RNA or ribonucleocomplexes is the most common system. This technique was used to edit CXCR4 and PD-1, knocking in new sequences to replace specific genetic letters in these proteins. The group was then able to sort the cells, using cell surface markers, to help identify successfully edited cells. Deep sequencing of a target site confirmed that knock-in genome

modifications had occurred with up to \approx 20% efficiency, which accounted for up to approximately one-third of total editing events. However, hard-to-transfect cells (stem cells, neurons, hematopoietic cells, etc.) require more efficient delivery systems such as those based on lentivirus (LVs), adenovirus (AdV) and adeno-associated virus (AAV).

CRISPRs have been used to cut 5 to 62 genes at once: pig cells have been engineered to inactivate all 62 Porcine Endogenous Retroviruses in the pig genome, which eliminated transinfection from the pig to human cells in culture. CRISPR's low cost compared to alternatives is widely seen as revolutionary. Selective engineered redirection of the CRISPR/Cas system was first demonstrated in 2012 in: (1) Immunization of industrially important bacteria, including some used in food production and large-scale fermentation. (2) Cellular or organism RNA-guided genome engineering. Proof of concept studies demonstrated examples both *in vitro* and *in vivo*. (3) Bacterial strain discrimination by comparison of spacer sequences.

Several variants of CRISPR/Cas9 allow gene activation or genome editing with an external trigger such as light or small molecules. These include photoactivatable CRISPR systems developed by fusing light-responsive protein partners with an activator domain and a dCas9 for gene activation, or fusing similar light responsive domains with two constructs of split-Cas9, or by incorporating caged unnatural amino acids into Cas9, or by modifying the guide RNAs with photocleavable complements for genome editing.

Methods to control genome editing with small molecules include an allosteric Cas9, with no detectable background editing, that will activate binding and cleavage upon the addition of 4-hydroxytamoxifen (4-HT), 4-HT responsive intein-linked Cas9s or a Cas9 that is 4-HT responsive when fused to four ERT2 domains. Intein-inducible split-Cas9 allows dimerization of Cas9 fragments and Rapamycin-inducible split-Cas9 system developed by fusing two constructs of split Cas9 with FRB and FKBP fragments. Furthermore, other studies have shown to induce transcription of Cas9 with a small molecule, doxycycline. Small molecules can also be used to improve Homology Directed Repair (HDR), often by inhibiting the Non-Homologous End Joining (NHEJ) pathway. These systems allow conditional control of CRISPR activity for improved precision, efficiency and spatiotemporal control.

Using dead versions of Cas9 (dCas9) eliminates CRISPR's DNA-cutting ability, while preserving its ability to target desirable sequences. Multiple groups added various regulatory factors to dCas9s, enabling them to turn almost any gene on or off or adjust its

level of activity. Like RNAi, CRISPR interference (CRISPRi) turns off genes in a reversible fashion by targeting, but not cutting a site. The targeted site is methylated, epigenetically modifying the gene. This modification inhibits transcription. Cas9 is an effective way of targeting and silencing specific genes at the DNA level. In bacteria, the presence of Cas9 alone is enough to block transcription. For mammalian applications, a section of protein is added. Its guide RNA targets regulatory DNA sequences called promoters that immediately precede the target gene.

Cas9 was used to carry synthetic transcription factors that activated specific human genes. The technique achieved a strong effect by targeting multiple CRISPR constructs to slightly different locations on the gene's promoter. In 2016 researchers demonstrated that CRISPR from an ordinary mouth bacterium could be used to edit RNA. The researchers searched databases containing hundreds of millions of genetic sequences for those that resembled CRISPR genes. They considered the fusobacteria *Leptotrichia shahii*. It had a group of genes that resembled CRISPR genes, but with important differences. When the researchers equipped other bacteria with these genes, which they called C2c2, they found that the organisms gained a novel defense.

Many viruses encode their genetic information in RNA rather than DNA that they repurpose to make new viruses. HIV and poliovirus are such viruses. Bacteria with C2c2 make molecules that can dismember RNA, destroying the virus. Tailoring these genes opened any RNA molecule to editing. CRISPR simplifies creation of animals for research that mimic disease or show what happens when a gene is knocked down or mutated. CRISPR may be used at the germline level to create animals where the gene is changed everywhere, or it may be targeted at non-germline cells.

CRISPR can be utilized to create human cellular models of disease. For instance, applied to human pluripotent stem cells CRISPR introduced targeted mutations in genes relevant to polycystic kidney disease (PKD) and focal segmental glomerulosclerosis (FSG). These CRISPR-modified pluripotent stem cells were subsequently grown into human kidney organoids that exhibited disease-specific phenotypes. Kidney organoids from stem cells with PKD populations formed large, translucent cyst structures from kidney tubules. Kidney organoids with mutations in a gene linked to FSG developed junctional defects between podocytes, the filtering cells affected in that disease. Importantly, these disease phenotypes were absent in control organoids of identical genetic background, but lacking the CRISPR modifications. A similar approach was taken to model long QT syndrome in cardiomyocytes derived from pluripotent

stem cells. These CRISPR-generated cellular models, with isogenic controls, provide a new way to study human disease and test drugs.

In 2003 evolutionary biologist Austin Burt envisioned attaching a gene that coded for a desired trait to selfish DNA elements that could copy themselves from one chromosome position to another. That would bias daughter cells to inherit it, quickly spreading it throughout a population. In 2015 a U.S. team used CRISPR to create a mutagenic chain reaction that drove a pigmentation trait in lab-grown *Drosophila* to the next generation with 97% efficiency. With another research group they created a gene drive in mosquitoes that spread genes that prevented the insects from harboring malaria parasites. Weeks later, the team reported a second drive with genes that rendered female mosquitoes infertile and could quickly wipe out a population.

CRISPR/Cas-based RNA-guided nucleases can be used to target virulence factors, genes encoding antibiotic resistance and other medically relevant sequences of interest. This technology thus represents a novel form of antimicrobial therapy and a strategy by which to manipulate bacterial populations. Some of the affected genes are tied to human diseases, including those involved in muscle differentiation, cancer, inflammation and fetal hemoglobin. Research suggests that CRISPR is an effective way to limit replication of multiple herpesviruses. It was able to eradicate viral DNA in the case of Epstein-Barr virus (EBV). Anti-herpesvirus CRISPRs have promising applications such as removing cancer-causing EBV from tumor cells, helping rid donated organs for immunocompromised patients of viral invaders, or preventing cold sore outbreaks and recurrent eye infections by blocking HSV-1 reactivation. As of August 2016, these were awaiting testing. CRISPR is being applied to develop tissue-based treatments for cancer and other diseases.

CRISPR may revive the concept of transplanting animal organs into people. Retroviruses present in animal genomes could harm transplant recipients. In 2015 a team eliminated 62 copies of a retrovirus's DNA from the pig genome. CRISPR may have applications in tissue engineering and regenerative medicine, such as by creating human blood vessels that lack expression of MHC class II proteins, which often cause transplant rejection.

In 2015, multiple studies attempted to systematically disable each individual human gene, in an attempt to identify which genes were essential to human biology. Between 1,600 and 1,800 genes passed this test—of the 20,000 or so known human genes. Such genes are more strongly activated, and unlikely to carry disabling mutations. They are more likely to have indispensable counterparts in other

species. They build proteins that unite to form larger collaborative complexes. The studies also catalogued the essential genes in four cancer-cell lines and identified genes that are expendable in healthy cells, but crucial in specific tumor types and drugs that could target these rogue genes. The specific functions of some 18% of the essential genes are unidentified. In one 2015 targeting experiment, disabling individual genes in groups of cells attempted to identify those involved in resistance to a melanoma drug. Each such gene manipulation is itself a separate drug, potentially opening the entire genome to CRISPR-based regulation.

In 2016-2017, a CRISPR/Cas-based approach to genetically engineering adult rodent brains *in vivo* was successfully demonstrated. Unenriched sequencing libraries often have abundant undesired sequences. Cas9 can specifically deplete the undesired sequences with double strand breakage with up to 99% efficiency and without significant off-target effects as seen with restriction enzymes. Treatment with Cas9 can deplete abundant rRNA while increasing pathogen sensitivity in RNA-seq libraries.

As of December 2014, patent rights to CRISPR were contested. Several companies formed to develop related drugs and research tools. As companies ramp up financing, doubts as to whether CRISPR can be quickly monetized were raised. In February 2017 the US patent office ruled on a patent interference case brought by University of California with respect to patents issued to the Broad Institute, and found that the Broad patents, with claims covering the application of CRISPR/cas9 in eukaryotic cells, were distinct from the inventions claimed by University of California. As of November 2013, SAGE Labs (now part of Horizon Discovery group) had exclusive rights from one of those companies to produce and sell genetically engineered rats and non-exclusive rights for mouse and rabbit models. By 2015, Thermo Fisher Scientific had licensed intellectual property from ToolGen to develop CRISPR reagent kits.

At least four labs in the US, labs in China and the UK, and a US biotechnology company called Ovascience announced plans or ongoing research to apply CRISPR to human embryos. Scientists, including a CRISPR co-inventor, urged a worldwide moratorium on applying CRISPR to the human germline, especially for clinical use.

In April 2015, Chinese scientists reported results of an attempt to alter the DNA of non-viable human embryos using CRISPR to correct a mutation that causes beta thalassemia, a lethal heritable disorder. The study had previously been rejected by both *Nature* and *Science* in part because of ethical concerns. The experiments resulted in changing only some genes, and had off-target effects on other genes. The

researchers stated that CRISPR is not ready for clinical application in reproductive medicine. In April 2016 Chinese scientists were reported to have made a second unsuccessful attempt to alter the DNA of non-viable human embryos using CRISPR - this time to alter the CCR5 gene to make the embryo HIV resistant.

In December 2015, an International Summit on Human Gene Editing took place in Washington under the guidance of David Baltimore. Members of national scientific academies of America, Britain and China discussed the ethics of germline modification. They agreed to support basic and clinical research under appropriate legal and ethical guidelines. A specific distinction was made between somatic cells, where the effects of edits are limited to a single individual, versus germline cells, where genome changes could be inherited by future generations. Heritable modifications could have unintended and far-reaching consequences for human evolution, genetically (e.g. gene/environment interactions) and culturally (e.g. Social Darwinism). Altering of gametocytes and embryos to generate inheritable changes in humans was defined to be irresponsible. The group agreed to initiate an international forum to address such concerns and harmonize regulations across countries.

Policy regulations for the CRISPR/cas9 system vary around the globe. In February 2016, British scientists were given permission by regulators to genetically modify human embryos by using CRISPR/Cas9 and related techniques. However, researchers were forbidden from implanting the embryos and the embryos were to be destroyed after seven days.

The US has an elaborate, interdepartmental regulatory system to evaluate new genetically modified foods and crops. For example, the Agriculture Risk Protection Act of 2000 gives the USDA the authority to oversee the detection, control, eradication, suppression, prevention, or retardation of the spread of plant pests or noxious weeds to protect the agriculture, environment and economy of the US. The act regulates any genetically modified organism that utilizes the genome of a predefined plant pest or any plant not previously categorized. In 2015, Yang successfully deactivated 16 specific genes in the white button mushroom. Since he had not added any foreign DNA to his organism, the mushroom could not be regulated under by the USDA under Section 340.2. Yang's white button mushroom was the first organism genetically modified with the Crispr/cas9 protein system to pass US regulation. In 2016, the USDA sponsored a committee to consider future regulatory policy for upcoming genetic modification techniques. With the help of the US National Academies of Sciences, Engineering and Medicine, special interests groups met on April 15 to contemplate the possible

advancements in genetic engineering within the next 5 years and potential policy regulations that would need to come into play. With the emergence of rogue genetic engineers employing the technology, the FDA has begun issuing new regulations.

The CRISPR/Cas9 system is composed of a short noncoding guide RNA (gRNA) that has two molecular components: a target-specific CRISPR RNA (crRNA) and an auxiliary trans-activating crRNA (tracrRNA). The gRNA unit guides the Cas9 protein to a specific genomic locus via base pairing between the crRNA sequence and the target sequence. In bacteria CRISPR loci are composed of a series of repeats separated by segments of exogenous DNA (of ~30 bp in length), the spacers. The repeat-spacer array is transcribed as a long precursor and processed within repeat sequences to generate small crRNAs that specify the target sequences (protospacers) cleaved by Cas9 protein, the nuclease component of CRISPR system. CRISPR spacers are then used to recognize and silence exogenous genetic elements at the DNA level. Essential for cleavage is a three-nucleotide sequence motif (NGG) immediately downstream on the 3' end of the target region, known as the protospacer-adjacent motif (PAM). The PAM is present in the target DNA, but not the crRNA that targets it. Upon binding to the target sequence, the Cas9 protein induces a specific double-strand break. Following DNA cleavage, the break is repaired by cellular repair machinery through non-homologous end joining (NHEJ) or homology-directed repair (HDR) mechanisms. With target specificity defined by a very short RNA-coding region, the CRISPR/Cas9 system greatly simplifies genome editing (Fisher, 2017).

CRISPR/Cas9 system greatly simplifies genome editing and has great promise in broad applications such as stem cell engineering, gene therapy, tissue and animal disease models, and engineering disease-resistant transgenic plants. CRISPRs were first discovered in archaea by Francisco Mojica at the University of Alicante in Spain. He proposed that CRISPRs serve as part of the bacterial immune system, defending against invading viruses. They consist of repeating sequences of genetic code, interrupted by spacer sequences – remnants of genetic code from past invaders. The system serves as a genetic memory that helps the cell detect and destroy invaders when they return. Mojica's theory was experimentally demonstrated in 2007 by Philippe Horvath.

The CRISPR/Cas9 system requires only the redesign of the crRNA to change target specificity. This contrasts with other genome editing tools, including zinc finger and TALENs, where redesign of the protein-DNA interface is required. Furthermore, CRISPR/Cas9 enables rapid genome-wide

interrogation of gene function by generating large gRNA libraries for genomic screening.

The CRISPR/Cas9 RNA-guided DNA endonuclease has contributed to an explosion of advances in the life sciences that have grown from the ability to edit genomes within living cells. Genome editing is an important technique in the life sciences. This technique changes the life body genes with high efficiency, high DNA sequence specificity, and little or no undesired byproducts. The natural genome-editing proteins evolved to achieve only partially related functions such as modulation of gene expression or protection from viral infection, and it is slow and small amount.

Early genome-editing efforts were enabled by the discovery that the endogenous cellular homologous recombination repair pathway could be used to replace a small portion of the genome of a living cell with an exogenous donor DNA sequence. To use this strategy for genome editing, the exogenous DNA sequence must have homology to the target genomic DNA site. Following transfection of the donor DNA, incorporation at the desired locus spontaneously occurs very inefficiently, at rates of less than 1%, depending on the cell type and cell state (Smithies et al., 1985; Thomas et al., 1986).

CRISPR/Cas9 is a new powerful technique for the gene editing target, and it can be used in medical treatment, cancer biology and agriculture, etc (Ma et al. 2015b). This article introduces recent research reports as references in the related studies (Ma, et al., 2016b). Use of CRISPR/Cas9 has escalated rapidly in recent years and is expected to become routine practice in molecular biology and related fields of research (Lowder, et al, 2017).

However, the gene editing technique as CRISPR/Cas9 has a big problem: it could create unexpected mutations all over the genome. Steven Salzberg found hundreds of mutations (in experimental mice) that weren't supposed to be there. The results contradicted earlier studies that showed CRISPR caused very few of these "off-target" mutations (Steven Salzberg, 2017). Off-target effects occur because of how CRISPR works. It has two parts. RNA makes a beeline for the site in a genome specified by the RNA's string of nucleotides, and an enzyme cuts the genome there. Trouble is, more than one site in a genome can have the same string of nucleotides. The CRISPR gene editing can cause hundreds of unintended mutations.

The human trials with CRISPR are already underway in China. As CRISPR research moves forward around the world, that could put the US in a tight condition. However, if China continues to experiment on human embryos using CRISPR/Cas9 gene editing technique while the US government does

not fund similar research, American researchers will be forced into the private sector if they want to compete. Once there, they could be beholden to corporate interests.

There are also many social, ethical and legal implications problems on using genome-editing technique in human germline, especially using the high efficient technique CRISPR/Cas.

Abbreviations of genome editing glossaries

- Cas = CRISPR-associated genes
- Cas9 = a CRISPR-associated protein containing 2 nuclease domains, that is programmed by small to cleave DNA
- Cpf1 = Prevotella and Francisella 1
- CRISPR = Clustered regularly interspaced short palindromic repeats
- crRNA = CRISPR RNA
- dCas9 = nuclease-deficient Cas9
- DSB = double-stranded break
- gRNA = guide RNA
- HDR = homology-directed repair
- HNH = an endonuclease domain named for characteristic histidine and asparagine residues
- indel = insertion and / or deletion
- NDA = deoxyribonucleic acid
- NHEJ = non-homologous end joining
- PAM = protospacer-adjacent motif
- PCR = polymerase chain reaction
- RNA = ribonucleic acid
- RuvC = an endonuclease domain named for an E Coli protein involved in DNA repair
- sgRNA = single guide RNA
- TALEN = transcription-activator like effector nuclease
- tracrRNA (trRNA) = trans-activating crRNA
- ZFN = zinc-finger nuclease

References

1. Alexis C. Komor, Ahmed H. Badran, David R. Liu. CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. *Cell*. Volume 168, Issues 1-2. DOI: <http://dx.doi.org/10.1016/j.cell.2016.10.044>. [http://www.cell.com/cell/fulltext/S0092-8674\(16\)31465-9](http://www.cell.com/cell/fulltext/S0092-8674(16)31465-9). 2017.
2. Fisher. <https://www.thermofisher.com/us/en/home/life-science/genome-editing/genome-editing-learning-center/crispr-cas9-technology-information.html>. 2017.
3. <http://www.yourgenome.org/facts/what-is-crispr-cas9>. 2017.
4. <https://en.wikipedia.org/wiki/Cas9>. 2017.
5. <https://en.wikipedia.org/wiki/CRISPR>. 2017.

6. Ma H, Young M, Yang Y. Cancer Biology Research Literatures. *Cancer Biology* 2015b;5(2):79-95. (ISSN: 2150-1041). <http://www.cancerbio.net>. doi:10.7537/marscbj070215.08. 8.
7. Ma H, Young M, Yang Y. Gene Knockout Research Literatures. *Rep Opinion* 2015;7(12):69-92. (ISSN: 1553-9873). <http://www.sciencepub.net/report>. doi:10.7537/marsroj071215.07. 7.
8. Ma H, Young M, Yang Y. Genome Editing. *Researcher* 2016;8(8):25-29. ISSN 1553-9865 (print); ISSN 2163-8950 (online). <http://www.sciencepub.net/researcher>. doi:10.7537/marsrsj080816.05. 5.
9. Ma H, Young M, Zhu Y, Yang Y, Zhu H. CRISPR and Stem Cell Research Literatures. *Stem Cell* 2016b;7(4):64-116. ISSN: 1945-4570 (print); ISSN: 1945-4732 (online). <http://www.sciencepub.net/stem>. doi:10.7537/marssej070416.13. 13.
10. Smithies, O., Gregg, R.G., Boggs, S.S., Koralewski, M.A., and Kucherlapati, R.S. Insertion of DNA sequences into the human chromosomal β -globin locus by homologous recombination. *Nature*. 1985; 317: 230–234.
11. Thomas, K.R., Folger, K.R., and Capecchi, M.R. High frequency targeting of genes to specific sites in the mammalian genome. *Cell*. 1986; 44: 419–428.
12. Steven Salzberg. CRISPR Gene Editing Controversy: Does It Really Cause Unexpected Mutations? <https://www.forbes.com/sites/stevensalzberg/2017/07/10/crispr-gene-editing-problems-does-it-really-cause-unexpected-mutations/#2468fd6a1a3>. 2017.
13. <https://www.statnews.com/2016/07/18/crispr-off-target-effects>. 2017.
14. <https://phys.org/news/2017-05-crispr-gene-hundreds-unintended-mutations.html>. 2017.
15. Levi Lowder, Aimee Malzahn, Yiping Qi. Rapid Construction of Multiplexed CRISPR-Cas9 Systems for Plant Genome Editing. https://link.springer.com/protocol/10.1007%2F978-1-4939-6859-6_25. 2017.
16. https://www.idtdna.com/pages/docs/default-source/synthetic-biology/crispr_faq_043015.pdf?sfvrsn=2. 2017.

8/25/2017