Beginner’s guide to CRISPR-Cas9-based gene editing

Shravanti Suresh (Iowa State University, USA)

In the last 20 years since its discovery, clustered regularly interspaced short palindromic repeats (CRISPR) has evolved from a mere bacterial immune system to a tool that can be programmed to achieve directed, efficient and sequence-specific modifications to the host genome. Its ease of use and minimum side effects to the cell lines are also a reflection of its ability to be extrapolated to medical environments, where it can be used for genetic modifications in the more complicated human genome. In research labs, CRISPR is now frequently used to achieve fast adjustments to host cell lines. It is used to knock in or knock out genes of interest and for generating models for genetic studies. This guide will help readers understand the basic requirements of a genome editing experiment using CRISPR-Cas components.

CRISPR, which refers to clustered regularly interspaced short palindromic repeats, was first reported in 1993 by Francisco Mojica. In 2000, he recognized disparate sequences followed by repeats and coined the term CRISPR, along with Ruud Jansen in 2002. He later found that these disparate sequences match snippets of genomes from bacteriophages, thus hypothesizing that CRISPR could be part of an adaptive immune system used by bacteria. In 2007, Phillippe Horvath and Rodolphe Barrangou, employees of a food company involved in cheese and yogurt production, showed that lactic acid bacteria could become resistant to bacteriophages when they incorporated sequences from the phage genomes. They also showed that a single protein, Cas9 from Streptococcus pyogenes, was required for resistance against these phages followed by their subsequent degradation. In the spring of 2011, during a meeting between Jennifer Doudna and Emmanuelle Charpentier at a conference in Puerto Rico, a collaboration was born. This scientific serendipity eventually led us to one of the fastest developing molecular biology tools that we have today, CRISPR-Cas based genome editing. Within 10 years, they won the Nobel Prize in Chemistry for their discoveries and insights into a bacterial immune system that has immense potential for modern research.

How the CRISPR-Cas systems work

The CRISPR array is encoded by a set of CRISPR-associated genes known as Cas genes. There are two major classes of CRISPR-Cas systems with six types and various subtypes. CRISPR immunity proceeds in three steps: adaptation, expression and interference.

During adaptation, Cas proteins recognize potential spacer sequences (protospacers) in the bacteriophage genome and excise them in order to incorporate them within their own. This recognition happens with the help of a conserved protospacer adjacent motif (PAM) sequence.

Upon integration, the spacers, along with a repeat in the CRISPR array, are transcribed or expressed to make precursor CRISPR RNAs (crRNAs), which are cleaved to form mature crRNAs. This guides the rest of the immune response. In some systems such as type II-A, during the expression stage, another short trans-activating crRNA (tracrRNA) is expressed. The tracrRNA is partially complementary to the repeat-derived segment in the pre-crRNA and together they form a dual crRNA-tracrRNA structure.

In the final step known as interference, these mature crRNAs, along with Cas proteins, look for complementarity in the invader genome and upon finding a match, a Cas nuclease cleaves the invader genome and eventually degrades it.

Each subtype of the CRISPR systems uses different proteins and mechanisms to recognize and cleave the target genome. Some types have been repurposed for biotechnological applications beyond aspects of native bacterial immunity. The targeting sequence in the crRNA can be changed in vitro for experimental purposes, allowing CRISPR-Cas machinery to target any region of interest and generate cuts. This makes CRISPR-Cas an effective pair of molecular scissors. When this occurs inside cells, these breaks are eventually fixed by DNA repair machinery which potentially results in an ‘edited’ genetic sequence.

Genome editing is performed by first generating a double stranded break (DSB) and, second, repairing that break while performing an edit. The creation of DSBs prompts the cell to repair the break in two ways as described in Figure 1.
A genome editing experiment requires the following components at the least:

- A host cell whose genome needs to be edited
- A guide RNA which is transcribed and targets a region of interest in the genome
- Cas proteins that are encoded within the cell or delivered to enable target binding and subsequent cleavage.

The following steps describe ways to design an efficient experiment and how to use these components to maximize the possibility of a successfully edited genome.

**Selecting the appropriate Cas protein**

The most commonly used protein in genome editing experiments is spCas9 found in *S. pyogenes*. spCas9 facilitates genome editing by generating DSBs. This cleavage is dependent on spCas9 binding to the 5’-NGG-3’ PAM sequence that is located downstream of the cleavage site. Depending on the Cas protein used for genome editing and the organism it is isolated from, the PAM will vary.

Despite its popularity, spCas9 has some limitations. It has the ability to recognize other PAMs such as 5’-NAG-3’ and 5’-NGA-3’. spCas9 can also bind and cleave sequences in the genome that have some similarity to the crRNA. These low-fidelity aspects of spCas9 result in off-target effects. spCas9 is also a rather large protein and delivery into cells can be a challenge. For this reason, scientists have discovered and engineered alternatives to spCas9 that can also be used in genome editing. Some examples of other commonly used Cas nucleases are Cas12a, Cas13a, IbgCas9, asCas9 and CasX. Other high-fidelity engineered variants including HypaCas9, eSpCas9 and Hi-FiCas9 offer many advantages over the traditional Cas9. These variants have a weakened DNA phosphate backbone interaction with the nuclease. This results in genome-wide specificity and undetectable off-target effects. However, each specific variant has certain improved characteristics which can benefit the process, depending on the type of experiment being performed.

**Guide RNA design**

In a genome editing experiment with spCas9, a guide RNA can be designed in two ways: a two-component crRNA plus a tracrRNA system or as a single guide RNA (sgRNA). In bacteria, a tracrRNA includes a segment at its 5’ end which is complementary to the repeat-derived segment of the pre-crRNA, allowing the two to hybridize. The guide region of the crRNA is typically 20 nucleotides long and is synthesized along with an
optimized tracrRNA length and joined by a linker loop to form one continuous molecule. sgRNAs are preferred as the separate crRNA and tracrRNA molecules may not anneal efficiently. Only the guide region in the sgRNA determines the region to be targeted in the genome and this flexibility is what makes CRISPR-based genome editing useful and easy. The DNA sequence encoding the guide RNA does not contain a PAM and this prevents it from being cut by the Cas proteins, though some applications of PAMs in homing guide RNAs have been developed. While designing guide RNAs, it is recommended to design three to four different guides to increase the chances of at least one of them binding efficiently and cleaving the target. While designing guide RNAs for a knock-in, the cut site should be as close to the knock-in site as possible. The recommended distance is less than 10 nucleotides between the two. However, if the knock-in is on just one allele, the cut site to knock-in site distance can be increased up to 20 nucleotides to increase the frequency of monoallelic edits to biallelic edits. It is important to keep in mind that guide designs must be optimized to reduce off-target binding and cleavage. There are many tools available for guide design that predict the on-target activity based on certain established scoring rules. Benchling, IDT (Integrated DNA Technologies), Broad Institute GPP, CasOFFinder, CRISPOR and Synthego have some useful resources that can help with this process.

**Delivery of the Cas protein and guide RNA into the cells**

Once the design is ready, the Cas reagents can be delivered into cells in these forms (Figure 2):

1) Naked DNA: The guide RNA sequences along with the Cas protein can be cloned into a plasmid which can then be delivered into cells and be selected for. Sometimes they are expressed on different plasmids but it’s easier when they are on the same plasmid and controlled by two different promoters. U6 is the most common promoter used for guide RNA expression and for spCas9, a stronger promoter such as T7, CMV or a cell-specific promoter is used. Delivery in the form of an mRNA is also another option. It is important to optimize the amount of DNA template based on each experiment and it can be delivered in these forms:

   a) Plasmid/double stranded DNA: These are suitable if the insert is long (>50 bp), especially if the goal is to replace an entire gene or insert a reporter gene. The template is in the form of a plasmid or double stranded fragments can be generated using PCR. The homology arms should be 50–800 bp in size to ensure efficient homology-directed repair. The disadvantage is that sometimes the plasmid DNA can integrate itself into off-target sites. Make sure to include a nuclear localization sequence for transcription within the cell. Usually, 400 ng of the plasmid donor (~5 kb) per 2.5 × 10^6 cells is sufficient for transfection.

   b) ssDNA: This method is useful when the knock-in fragment is less than 50 bp. The homology arms are recommended to be 50–80 bp long (50 bp long for single nucleotide changes). It is recommended to start with at least 60 pmols of ssDNA donor template per 1 × 10^6 to 2.5 × 10^6 host cells for high-efficiency knock-ins. These should be end-modified to be protected from cellular nuclease.

2) Viral transduction methods: Different types of viral vectors are used to transduce DNA/RNA into the host cells. To do so, plasmids containing the viral genes and the CRISPR-Cas components are introduced into a packaging cell line (e.g., 293 T cells) from which viral particles are harvested. These are later introduced to a host cell line. Examples of such vectors are:

   a) Adeno-associated viruses: These non-pathogenic single stranded viruses can be modified by removing all their viral genes and inserting the sequence of interest containing inverted terminal repeats. They have a natural ability to induce homologous recombination; however, the drawbacks are that the size of the insert is limited to 4.5 kb and this method is generally more expensive. The advantage, however, is that since AAVs do not trigger an immune response and do not integrate within the genome, allowing less chances for insertional mutagenesis.

   b) Lentiviruses: These belong to the family of retroviruses and integrate into the genome of the host cells by penetrating through the nuclear envelope. This makes them highly effective in non-dividing cells but also increases their chances of disrupting essential genes.

3) Riboprotein complex: The Cas protein-guide RNA complex can be delivered into cells via various methods such as lipofection, electroporation, nucleofection and microinjection. For a preformed RNP complex, the transfection can be done directly into the nucleus or via microinjection. The type of transfection method used will depend on whether there is a need for permanent expression of the CRISPR-Cas components or if a transient expression would suffice.

**Tips**

- Sometimes spCas9 can re-cut the DNA after it is inserted within the genome. In such cases, designing the donor insert with a PAM containing a silent mutation can make the PAM non-functional upon insertion and Cas9 will no longer bind to it and re-cut after insertion.
Some studies suggest making asymmetric homology arms or protecting the DNA ends with phosphorothioate to improve efficiency.

While performing a genome editing experiment, it is important to use marker genes to determine if the foreign genetic sequences have been inserted successfully into the organism of interest. A common screenable marker includes GFP, which is cloned into the plasmid along with the Cas protein and guide RNA. If fluorescence is detected after transfection, that is usually a good sign though further validation with functional assays or genotyping needs to be done. While this will suffice for transient expression, to stably incorporate the transgene (GFP) in the cell line, multiple rounds of single-cell isolation and expansion are required. For stable transfections, a marker such as antibiotic resistance marker may also be used. Stable transfections are usually more efficient and thus may not require markers.

Optimization parameters also include concentration of sgRNA, sgRNA:Cas protein ratio (it is recommended to have a ratio between 3:1 and 9:1 if spCas9 is used and depending on the experiment), reagent volume, cell number, recovery media and the electroporation/nucleoporation settings.

Data analysis with CRISPR

Once the genome is edited, it needs to be validated. Regardless of the design of the experiment or the method of repair, the first step is always to PCR-amplify the edited region. This amplified region can then be subject to Sanger sequencing or high-throughput sequencing.
After the sequencing results arrive, remember to compare it to a control sequence to determine the presence of off-target mutations, if any, and to account for any other indels in the target region. Most programs used for CRISPR data analysis provide scores that can be used as an indication of a successful genome editing experiment. Once the edit is bias confirmed, you are now ready to do some exciting science!

Recent applications of CRISPR

CRISPR-Cas-based diagnostic testing for the SARS-CoV-2 virus during the COVID-19 pandemic is a recent example of its versatility. Genome editing using CRISPR is used to generate animal models with ease to bridge the gap between laboratory-based validation and human clinical trials. A catalytically dead spCas9 can activate and silence genes in modelling experiments. CRISPR-Cas9 is also being used in cell therapy applications to correct defective phenotypes. While many studies show that it can be implemented in vivo, we need more information about its mutational competency and DNA repair pathways to completely determine the efficacy of CRISPR-based therapy. Nevertheless, targeting CRISPR-Cas9 riboproteins to blood, liver, eye and muscle cells is being extensively studied and will be achievable in the near future. CRISPR also has agricultural applications, including genetically modifying crops to endure biotic and abiotic stresses, to enhance their nutritional capabilities and to combat plant pathogens. Apart from these applications, CRISPR has enormous potential to positively impact human life and the environment, if practiced and applied within the regulated confines of scientific experimentation, keeping in mind its legal, social and ethical implications.

Further Reading

- CRISPR-Cas biochemistry toolkit – https://doi.org/10.1016/j.molcel.2017.09.007
- For guide RNA design – https://zlab.bio/guide-design-resources
- Choosing an appropriate Cas nuclease – https://www.synthego.com/guide/how-to-use-crispr/cas9-nuclease-variants

Shravanti Suresh is a PhD candidate in the Department of Biochemistry, Biophysics and Molecular Biology at Iowa State University. She received her bachelor's degree in biotechnology from Anna University in Chennai, India. Her research focuses on investigating the mechanisms of adaptation in type I CRISPR-Cas systems in Sashital Lab. She is extremely interested in communicating science to a non-scientific audience. Email: shravantisuresh@gmail.com.