The Chemistry of CRISPR: Editing the Code of Life/ CRISPR-Cas9: Biology and Technology of Genome Editing*

Nobel Lecture, December 8, 2020 by
Jennifer Doudna
University of California, Berkeley, CA, USA.

I THOUGHT I WOULD BEGIN BY TELLING YOU about the origin of the ideas around CRISPR. The work began at least two decades ago with research in microbiology laboratories that showed bacteria might have an adaptive immune system, a way to provide protection against viral infection. This adaptive immune system would protect cells from bacteriophages – viruses that inject their DNA into bacteria – through a recording system in the bacterial genome that came to be called CRISPR, which stands for Clustered Regularly Interspaced Short Palindromic Repeats.

---

* I'd like to begin by thanking the Royal Swedish Academy of Sciences, the Nobel Prize Committee in Chemistry; my family, including my spouse, Jamie Cate, our son, Andy, and my sisters, Ellen and Sarah Doudna; my friends, colleagues, and, of course, my former and current lab members, about whose research I will be speaking today. It's a wonderful honor to have this opportunity to share with you the science that we've done over the last few years, and to discuss the extraordinary opportunities and exciting advances that are happening right now with CRISPR-Cas9 as a genome editing technology.

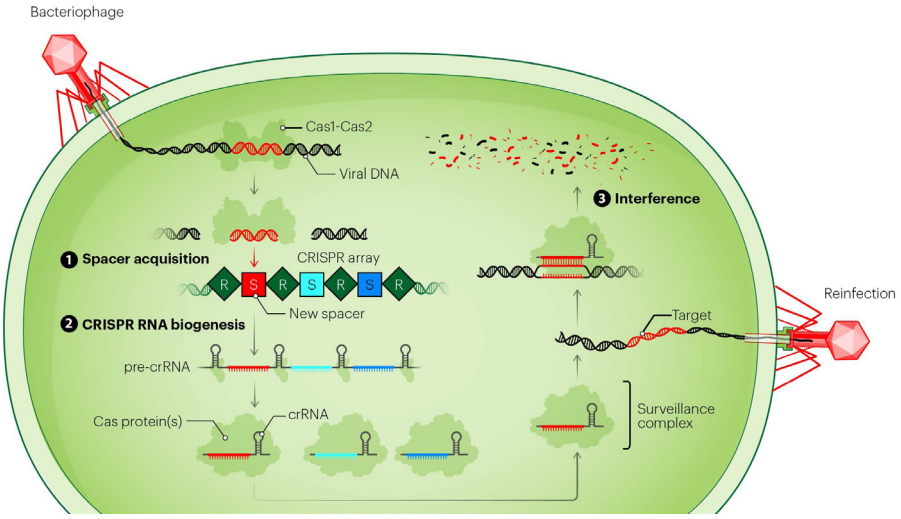

Figure. 1. The illustration above shows the way the CRISPR system is thought to act in bacteria. It represents experimental results that were done in the field early on using bioinformatics and molecular genetics to understand the function of the CRISPR pathway.

We imagine that in nature, when a viral bacteriophage injects its DNA into a bacterial cell, the cell can integrate small pieces of that foreign DNA into its genome at a locus or site called CRISPR, which consists of alternating sequences called repeats and spacers. A CRISPR RNA molecule (cRNA) makes a copy of the viral DNA that is passed on to CRISPR-associated or Cas proteins, which are enzymes that use this genetic information to detect and destroy the viral DNA by cleaving it. Through the CRISPR RNA biogenesis process, the bacterial cell also acquires immunity from similar invasions in the future.

Over time, the CRISPR immune system has evolved and diversified into many different forms, but today I'm actually going to talk in particular about one type of CRISPR-Cas system that uses the protein Cas9.

Back in 2011, we had the good fortune to begin collaborating with Emmanuelle Charpentier and her student, Krzysztof Chylinski. This collaboration launched a wonderful opportunity to answer what was at the time a very interesting and intriguing question in the CRISPR field: What is the function of the Cas9 protein?

We were fascinated by this question because Cas9 had been implicated in the protection of cells against invasive DNA, particularly in *streptococcus pyogenes*, a kind of bacteria Emmanuelle's group was studying that infects humans. These bacteria have a Cas9 protein encoded in their CRISPR system that was implicated in protecting the cells from viral infection. But the question was how? In our collaboration, we addressed

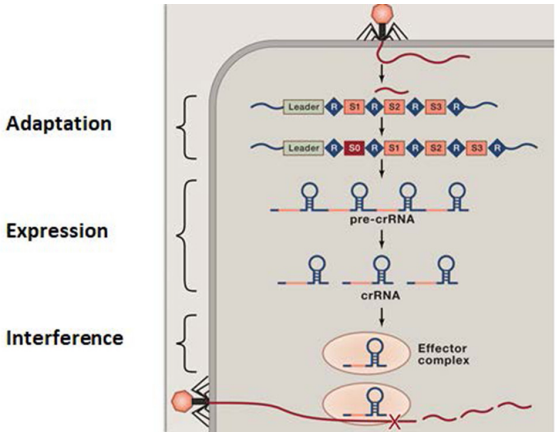

Figure 2. The illustration above shows the three steps in CRISPR acquired immunity that include adaptation, expression and interference. In the research that I did in my laboratory with a number of my former students, beginning with Blake Wiedenheft and then Rachel Haurwitz, and folks that came after them, we began investigating in particular, interference, the third step in the CRISPR pathway. Interference involves an RNA-guided detection of viral DNA.

this question by doing biochemical experiments that involved working withpurified CRISPR-Cas9 protein and the RNA that guides it to target DNA sequences and cells.

From research done by Martin Jínek, a former postdoc in my laboratory, and Krzysztof in Emmanuelle's laboratory, we found that CRISPR-Cas9 in nature is a dual RNA-guided protein.

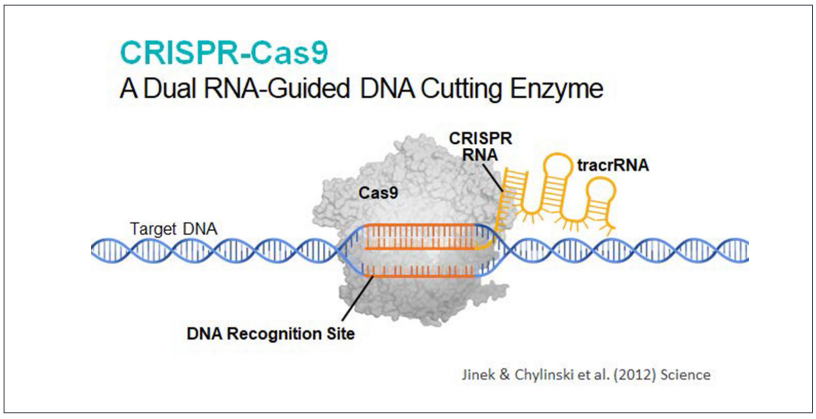

Figure 3. CRISPR-Cas9 A dual RNA-guided DNA cutting enzyme.

The Cas9 protein uses a CRISPR-RNA molecule to direct it to a target sequence of viral DNA that matches the CRISPR-RNA sequence. The process also requires a second RNA molecule called tracer – or tracrRNA – that interacts with the CRISPR RNA to guide Cas9 to target DNA sequences. Once it reaches the target, Cas9 is able to make a cut in the double helix of the viral DNA.

One of the wonderful aspects of CRISPR and this project with Emmanuelle's group is that we reached a point in our research where what began as a curiosity-driven investigation morphed into a project that had much broader implications.

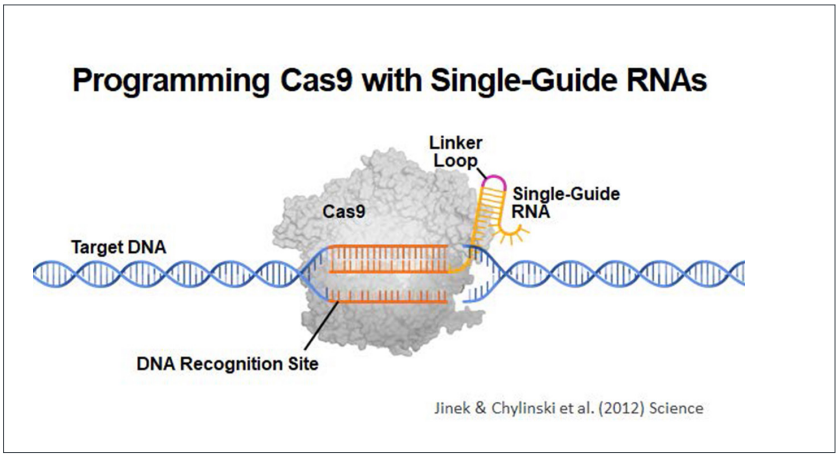

Figure. 4. Programming Cas9 with Single-Guide RNAs.

Once we understood how nature uses a dual-RNA system to guide Cas9 to target DNA sequences, it became possible to engineer the dual-guide RNA as a single-guide RNA molecule that includes both the targeting information and the structural requirement for assembly with the Cas9 protein. This single-guide RNA can then be used to program Cas9, directing it to cleave double-stranded DNA at a desired sequence. To accomplish this, we take advantage of the actual targeting information in the CRISPR-RNA molecule and a short sequence of DNA in the virus called "PAM" for protospacer adjacent motif. The presence of a PAM sequence adjacent to the target sequence in the viral DNA activates the RNA-guided Cas9 to cleave the target.

I want to show you a key experiment that Martin did to test this idea of using a single-guide RNA to target Cas9 proteins to cleave specific sequences of DNA. The experiment was to design guide RNAs that would recognize several different sequences in a plasmid DNA, a circular piece of double-stranded DNA that we could purify in the laboratory.

What you're seeing in the above image are five different sites in the

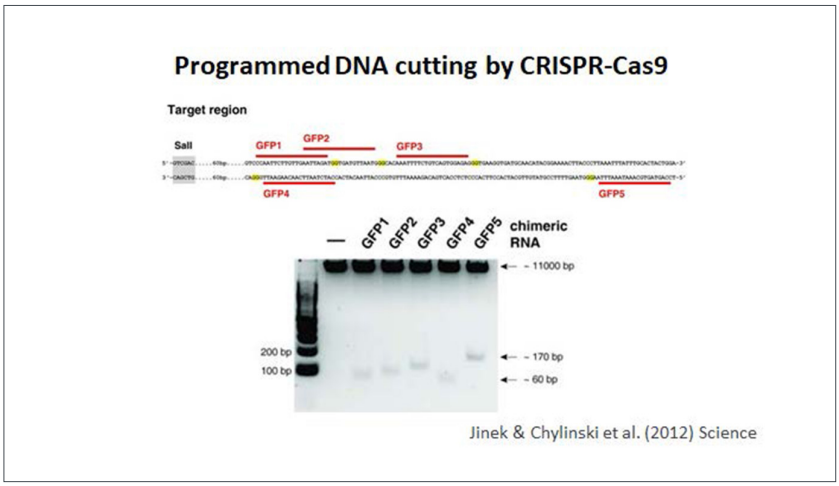

Figure. 5. Programmed DNA cutting by CRISPS-Cas9.

plasmid DNA – marked in red – that were chosen as target sequences for single-guide RNAs we produced in the laboratory. Martin added those single-guide RNAs to a purified Cas9 protein and incubated them together with the plasmid DNA molecule in a laboratory test tube. To analyze the results of that experiment, he separated the different cleaved DNA products from one another in an agarose gel system shown in the image. What you can see in each lane of this gel system is that depending on where the guide-RNA was directed to interact with the plasmid DNA, Cas9 would generate a cut. By also cutting the plasmid at a separate place so the two double-stranded breaks were introduced into the plasmid at one time, we could release these cut fragments of DNA into the gel system. As you can see, each cut fragment of plasmid DNA migrated to a different position based on the size of the fragment.

I have to say that on the day Martin did this experiment and got these results, we were just incredibly excited. It was the pure joy of discovery at recognizing that we not only understood how the Cas9 bacterial protein functions, but we had also actually figured out how to engineer it as a simple two component system for directing DNA double-stranded cutting. Why was that so exciting? Well, it was, of course, interesting to know that we could harness our knowledge in this way and engineer the protein to have this desired cleavage capability. But, in addition, it also allowed us to imagine how CRISPR-Cas9 could actually be harnessed as a technology for something quite different in eukaryotic cells: cells like plant, animal and human cells, all of which treat double-strand breaks in DNA differently than the way they're treated in bacteria. I'll show this in the next couple of images.

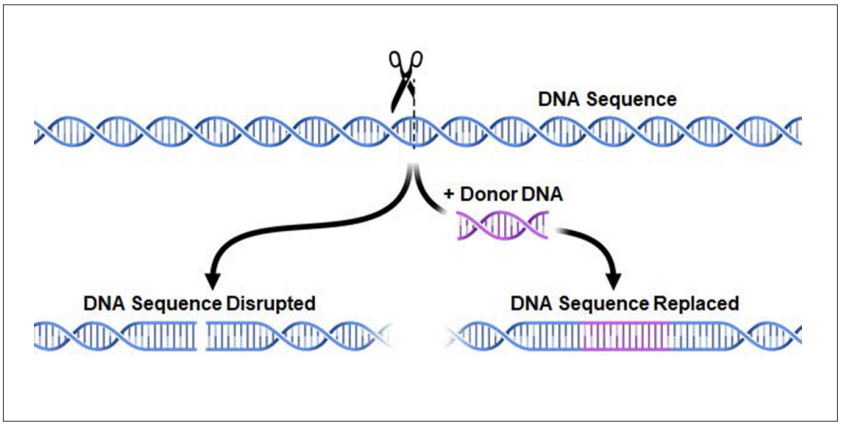

Figure. 6. ???

Typically, in a eukaryotic cell, when the cell receives or detects a double-stranded break in its DNA, the break is detected and repaired before it can cause cell death. The repair pathways involve either a non-homologous end-joining event (left side of the image) that introduces a small disruption to the DNA sequence, or there can be an integration of DNA that has homology to the sequence of DNA flanking the double-stranded break (right side of the image). In the latter case, a new piece of genetic information is incorporated into the genome at the site of the original break.

This research had been done over the previous couple of decades before Emmanuelle and I did our work with Cas9. We recognized that the ability of Cas9 to introduce a double-stranded break into DNA at a desired position could allow scientists to introduce double-stranded breaks into a genome that could trigger the kind of repair shown in this image.

We then imagined a system where the Cas9 protein could be directed to enter the nucleus of a eukaryotic cell and – directed by its guide-RNA – search the cell's DNA for a 20-base pair sequence that matches the 20-base pair sequence of the RNA guide. When that match occurs, we now understand that Cas9 is able to unwind the DNA and generate a precise double-stranded break by cleaving each strand of DNA. The broken ends are handed off to repair enzymes in eukaryotic cells that lead to DNA repair. In the process of repairing the broken DNA, there is the opportunity to introduce a change to the genome at a precise place. This is really the definition of genome engineering.

In the next few images, I want to share a few things we've learned over the last few years about how Cas9 is able to achieve this kind of editing in the genomes of cells by triggering double-stranded breaks.

I'll start with showing a molecular model of the Cas9 protein.

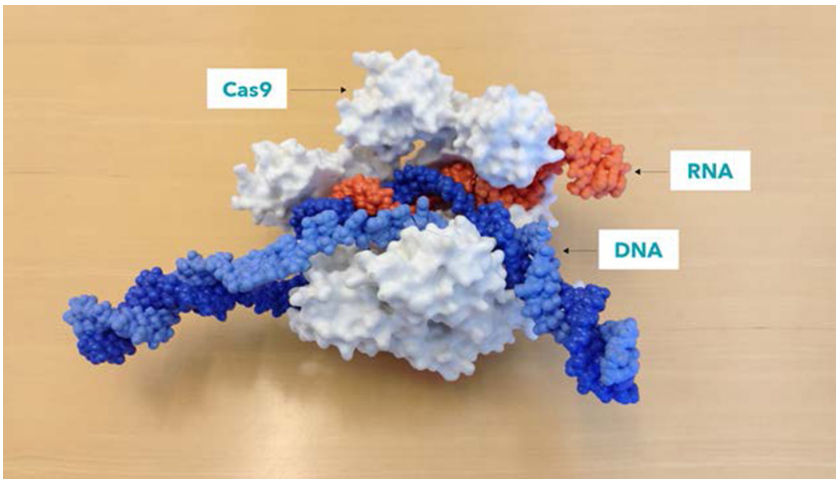

Figure. 7. A molecular model of the Cas9 protein.

This model is based on a crystallographic structure that was solved originally by several different laboratories; our own, the lab of Martin Jínek, the lab of Osamu Nureki, and subsequently many others who have contributed to understanding the actual molecular basis for Cas9 functions. In this model, Cas9 is the white protein holding on to its orange guide RNA and a blue double-helical DNA molecule. I'd like to point out that in this structure, we can see the mechanism by which Cas9 uses a guide RNA molecule to interact with DNA at a precise position.

The orange and blue helix inside the Cas9 protein represents the interaction between the guide RNA and a single strand of the target DNA molecule. This is the actual way that these proteins are able to find and hold onto DNA prior to cleaving it.

The next thing we learned in studying how this cleavage event actually works is that Cas9 is a highly dynamic protein. It has to be in order to be able to handle DNA and unwind a double helix the way we know it does. It has to be able to move, and, through a series of chemical experiments that allowed us to detect motions in different parts of the protein, we discovered that Cas9 also has to be able to undergo a large conformational change as it holds on to DNA and catalyzes cutting.

The Cas9 protein morphs to the structure it forms when it binds to the guide RNA. This is the Cas9 structure that is able to search the cell, looking for a matching sequence of DNA. When that match occurs, there's an additional structural change in the protein that accommodates the DNA molecule – the RNA-DNA helix forming inside Cas9. Finally, this yellow part of Cas9 swings into position so that it can cleave the DNA strand that is attached to the guide RNA. This is a very important aspect of the chemistry by which Cas9 cuts DNA because it provides a mechanism for

sensing the interaction between the guide RNA and the target DNA, and ensuring the accuracy of Cas9's cutting mechanism.

In work that was done just over the last few years, Fuguo Jiang, a former postdoc in my laboratory, and sadly now deceased, did a series of very exciting structural experiments to reveal the shape of the Cas9 protein when it's engaged on a full-length DNA molecule.

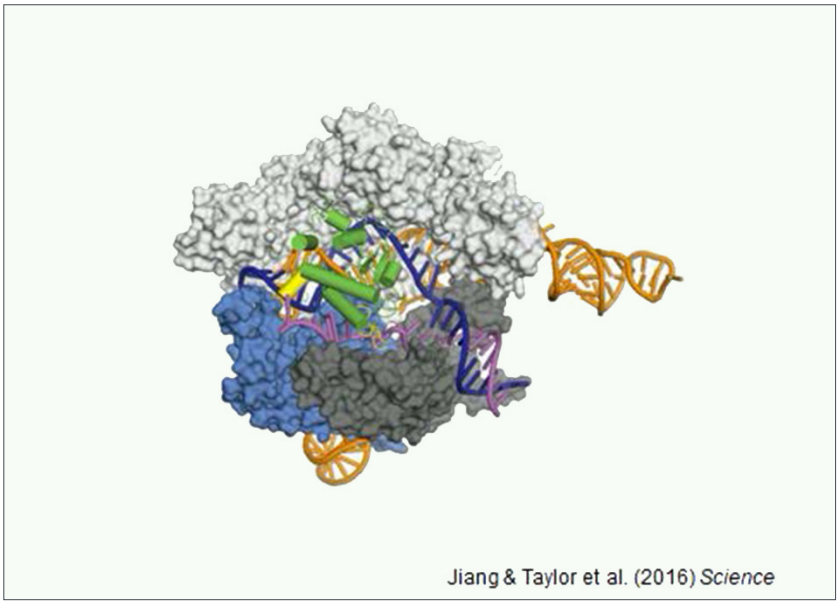

Figure. 8. In the image above, the guide RNA is again shown in orange; the DNA strands are in blue and magenta. The DNA is held open by the Cas9 protein and allowed to interact with the cleavage sites in the enzyme for precise double-stranded cutting of both strands of the DNA. As shown here in green, one of the cutting parts of the enzyme swings into position to perform the catalytic chemistry required to cleave the DNA strand.

I want to also point out that in addition to these conformational changes that happen in the Cas9 protein itself, we also now understand this protein is quite dynamic in the way that it interacts with long pieces of DNA – for example, chromosomes and cells. It has to be able to move very quickly, along the length of DNA, searching for a sequence that will bear a complementary match to the guide RNA. How does that work? It seemed like an extraordinary capability to us initially.

However, in experiments that we did in the laboratory over several years with a number of former lab members and collaborators, we came up with the model shown in this image that I think is consistent with current data.

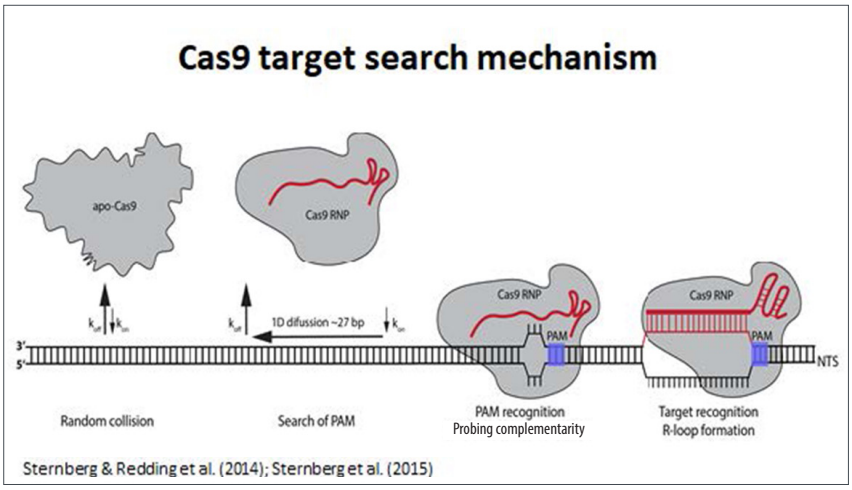

Figure. 9. Cas9 target search mechanism.

The model suggests that the Cas9 protein has the ability to bind and release DNA very quickly, allowing it to search through large, really vast stretches of DNA quite fast. This model also allows interrogation of the sequence that's being searched for a match with the guide RNA, which is shown here in red. You can see that our data suggests that when Cas9 protein with its guide RNA interacts with DNA it begins to pry apart the two strands of the double helix. This allows the protein to determine if there is a complementary sequence for binding to the CRISPR-RNA. If there is, then our data suggests that the strands of the DNA continue to melt apart, allowing the RNA-DNA helix to form inside the Cas9 protein. If that helix is perfect or close to perfect, then the enzyme is triggered to cleave DNA.

This is quite an amazing mechanism that clearly allows the CRISPR immune system to search the cell very quickly, looking for viral DNAs to destroy. But in eukaryotic cells, this mechanism is equally effective at triggering double-stranded breaks that can be repaired and triggering changes in the format of genome editing that we now understand can be effectively catalyzed by Cas9.

In the next part of the talk, I want to turn my attention to where CRISPR technology is going. There's a lot that one could say here so I'm just going to hit on a few of the highlights.

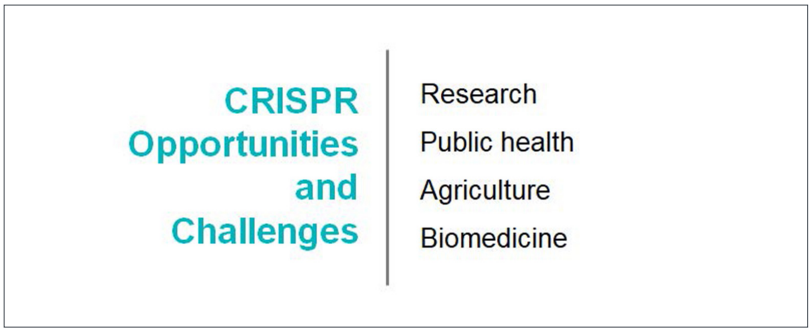

Figure. 10. CRISPR opportunities and challenges.

First of all, genome editing extends across all of biology. It can be used for fundamental research as well as for exciting applications in public health, agriculture, and biomedicine.

I think it's also very important to point out that genome editing can be conducted in many different kinds of cells, and fundamentally in the two kinds of cells cited, somatic cells and germ cells.

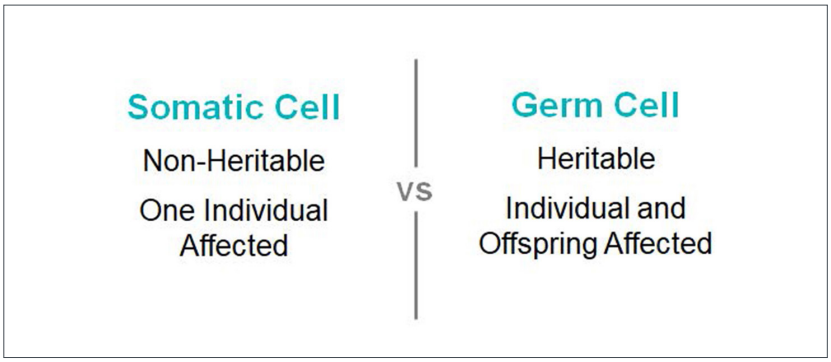

Figure. 11. Somatic cell vs. germ cell.

A somatic cell is fully differentiated and does not have the ability to create a new organism. A germ cell, such as a sperm cell, an egg cell, or cells in an early embryo, is pluripotent, which means it is able to differentiate into many different cell types as an organism is forming.

If genome edits are introduced into a somatic cell, those DNA changes are not heritable; they affect only one cell or one tissue type, or one individual organism. But if genome edits are introduced into a germ cell, the DNA changes have the potential to be heritable and to introduce DNA changes that become part of not only an individual but all of that individual's progeny.

Heritability makes genome editing of germ cells a very powerful tool when we think about using it in plants or using it to create better animal models of human diseases, for example, as has been done using CRISPR-Cas9 in mice and rats. It's very different when we think about the enormous ethical and societal issues raised by the possibility of using germline editing in humans. I won't say too much more about that, but a very active area of my own work over the last few years is to think about the responsible use of CRISPR-Cas9. In particular, I want to ensure that transparency and careful thought goes into work in which CRISPR-Cas9 is applied to the human germline.

In somatic cell genome editing, however, I think there are extraordinary and exciting opportunities that we will be developing in the near future. One of these opportunities is no longer a "potential," but has actually been realized.

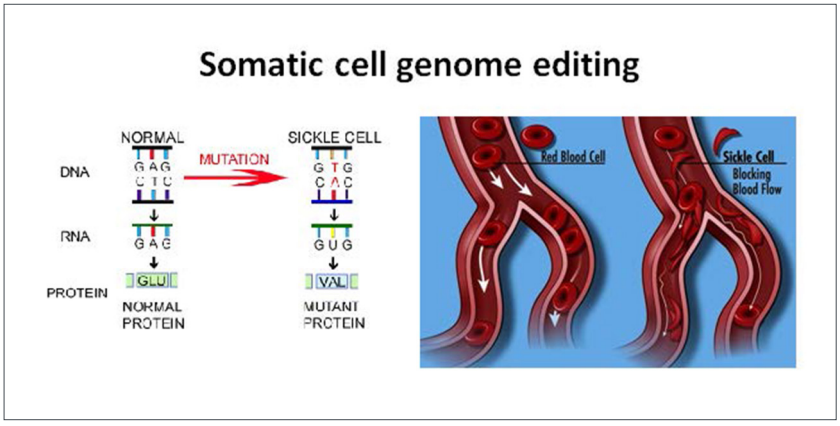

Figure. 12. Somatic cell genome editing.

CRISPR-Cas9 technology has been shown to be a safe and effective way to treat sickle cell anemia in a patient by correcting the well-defined mutation that causes the disease. I think this is a very exciting way to imagine how CRISPR-Cas9 technology will impact human health in the future.

With regards to germline editing, I certainly imagine that we will see increasing applications in germ cells, including in the human germline. That has to be managed very carefully. I'm pleased that there's been an active international effort to control the use of CRISPR-Cas9 and certainly to encourage transparency.

Fig 13. What about human germline editing?

Ethics
HHGE Consensus
Study Report

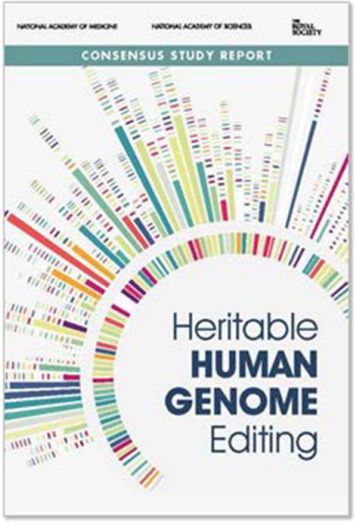

Figure. 14. HHGE consensus study report.

A recent joint report from the National Academy of Medicine, the National Academy of Sciences and the Royal Society discusses the science and the technology around human germline editing and establishes criteria for using CRISPR-Cas9 in the human germline in the future.

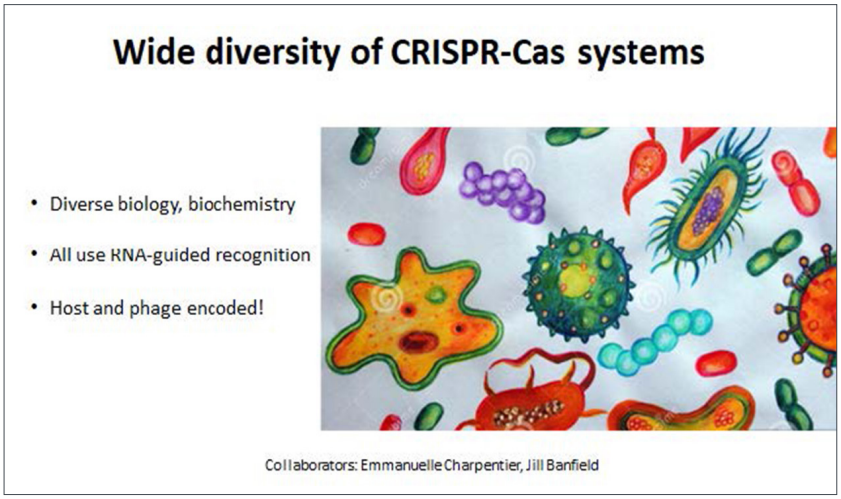

Figure. 15. Wide diversity of CRISPR-Cas systems.

In the last part of my talk, I'd like to turn to where CRISPR technology is headed in the future.

One of the aspects of this work that's fascinating to me is the incredible diversity of CRISPR systems in nature. This diversity continues to drive the field in terms of fundamental biology; understanding what these systems do in their natural settings and microbes; and how they may be harnessed as technologies and other organisms, as was the case for CRISPR-Cas9.

I also wanted to mention briefly the effort to investigate new CRISPR-Cas systems. One of the recent findings we've had with our collaborator, Jillian Banfield at the University of California, Berkeley, is the discovery that bacteriophages, the viruses that bacteria use CRISPR to protect themselves from, in fact can also carry around their own CRISPR-Cas systems.

As shown in the image next page, one example, a protein we named CRISPR-CasØ, is entirely phage-encoded. It's a tiny protein but nonetheless has the RNA-guided, DNA-cutting capabilities that we discovered originally in CRISPR-Cas9. This is a fascinating example of nature's diversity and provides opportunities for future applications of CRISPR-CasØ. For example, there are cases in which one could benefit from having a very tiny protein with a small gene that could be more easily delivered into eukaryotic cells than CRISPR-Cas9.

We've also been very interested in the biochemical activities of other CRISPR-Cas proteins. I'll mention here very briefly research done originally by Alexandra East-Seletsky, a former graduate student in my laboratory, and Mitch O'Connell working in partnership with her. Alexandra

and Mitch discovered that Cas13, a class of CRISPR proteins that are naturally occurring RNA-targeting enzymes, has a biochemical activity that could be harnessed as an RNA detection mechanism.

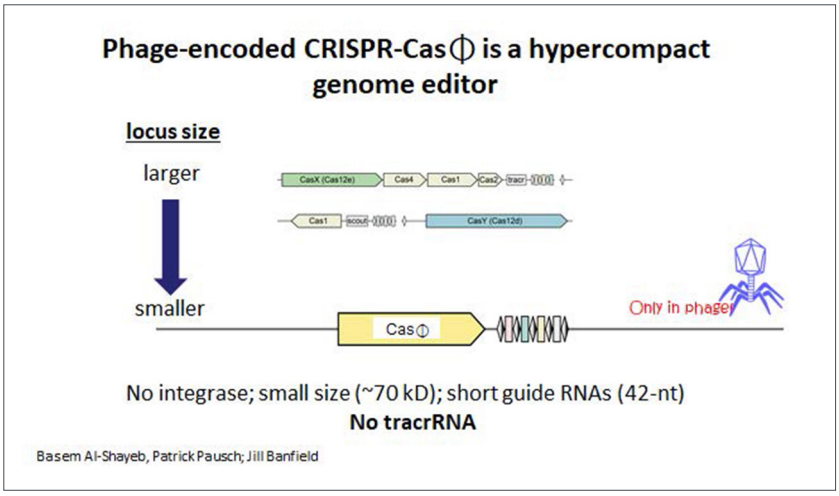

Figure. 16. Phage-encoded CRISPR-CasØ is a hypercompact genome editor.

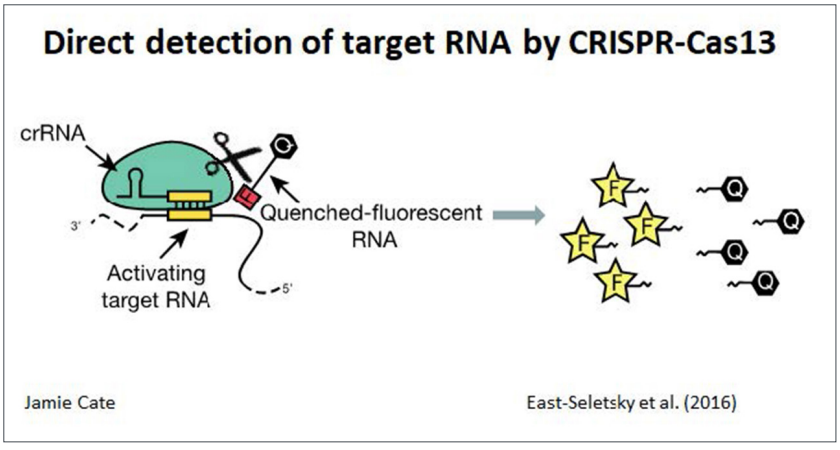

Figure. 17. Direct detection of target RNA by CRIPR-Cas13.

The experiment originally performed by Alexandra and suggested by Jamie Cate is shown in the above image. The idea was to put fluorophores on to small pieces of RNA that were cleaved upon Cas13's detection of an RNA sequence using its guide-RNA.

The CRISPR-Cas13 system turned out to be highly effective at detecting RNA molecules all the way down to about picomolar levels.

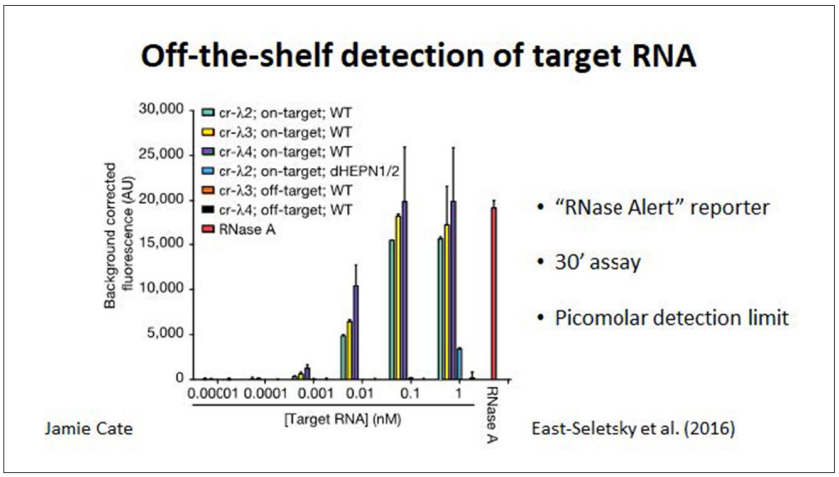

Figure. 18. Off-the-shelf detection of target RNA.

This CRISPR-Cas13 experiment was carried out with an off-the-shelf kit called RNase Alert. The results, shown in the image above, triggered our interest in the biochemical activities of other families of CRISPR-Cas proteins in addition to Cas9.

In research done a couple of years ago, Janice Chen, a student in my lab at the time, showed that the family of Cas proteins called Cas12 also has the ability to cleave single-stranded molecules. In this case, Cas12 cleaves single-stranded DNA upon recognition of a double-stranded DNA target.

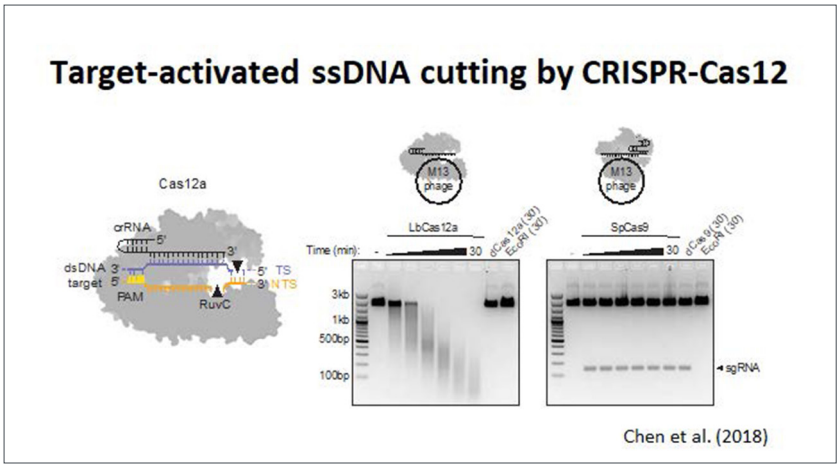

Figure. 19. Target-activated ssDNA cutting by CRISPR-Cas12.

In the experiment illustrated above, Janice showed how the Cas12a protein, upon recognition of a double-stranded DNA, is able to cleave single-stranded molecules of DNA. In the cEnter panel of this image, we see very rapid degradation of a circular single-stranded DNA molecule, which is a biochemical activity we do not detect for Cas9 as shown in the panel on the right.

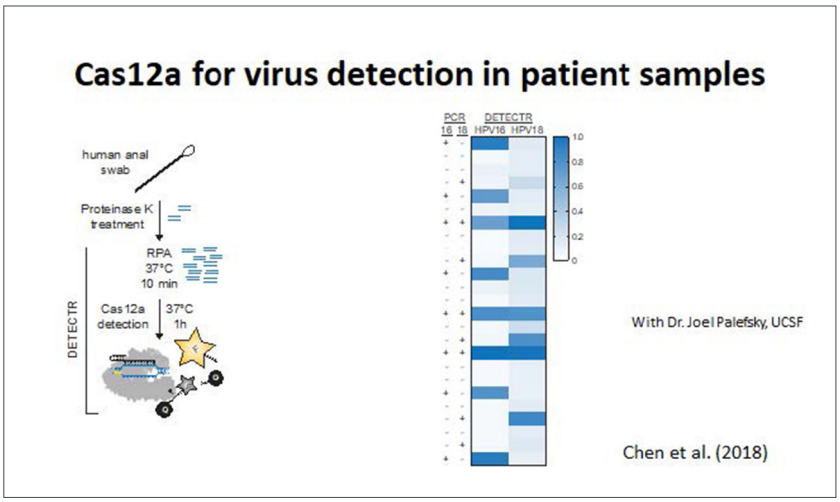

Figure. 20. Cas12a for virus detection in patient samples.

Working with a colleague, Joel Palefsky at University of California, San Francisco, we were able to use Cas12a activity to detect the human papillomavirus in patient samples, and even to distinguish between two different strains of HPV as shown in the image above.

This study told us that not only can CRISPR-Cas proteins be useful for detection; they can also be useful for specificity in figuring out what type of viral signal might be present in a patient's sample.

In the current SARS-CoV-2 pandemic that has become a worldwide human health issue, we and others have been using a detection technique based on CRISPR-Cas13a to very quickly identify SARS-CoV-2 infections in patient saliva or in patient nasal swabs. Ultimately, as shown in the image below, we hope to develop a rapid Covid diagnostic technique based on CRISPR-Cas13a that would allow people to test themselves and record the results on their cell phones. This would enable people everywhere to screen themselves against the Covid virus. The programmability of CRISPR-Cas proteins, which is a fundamental property of their biology, and their ability to be harnessed as technology should also help us prepare for future pandemics.

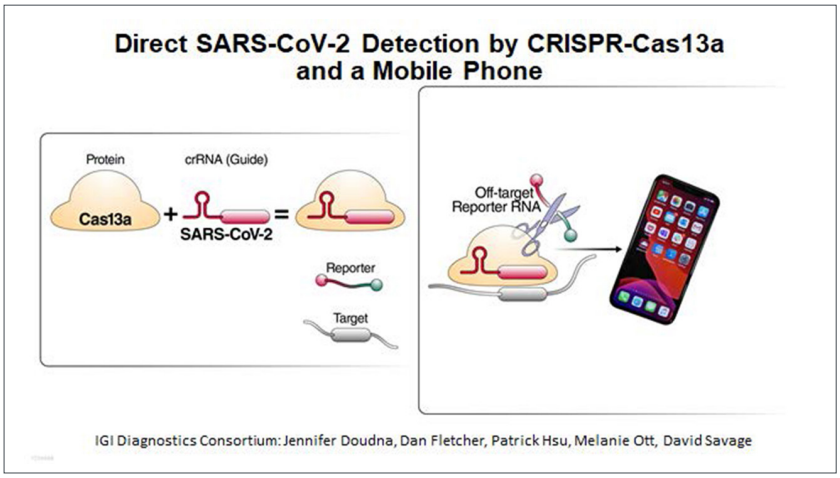

Figure. 21. Direct SARS-CoV-2 detection by CRISPR-Cas13a and a mobile phone.

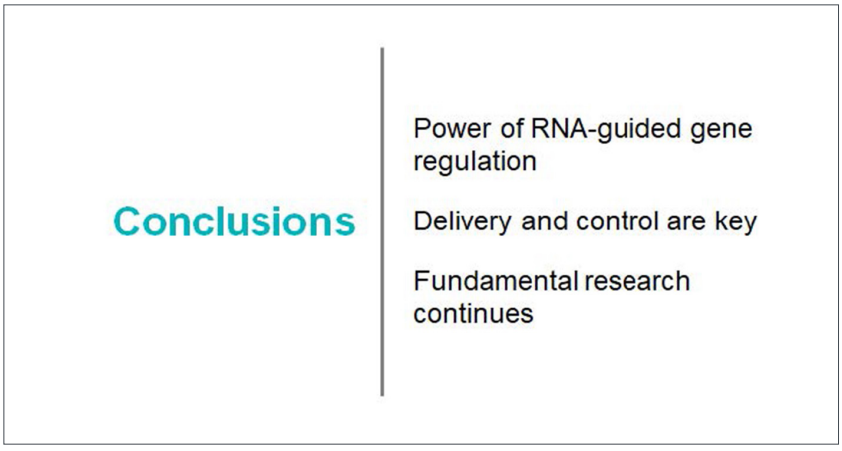

Figure. 22. Conclusions.

I'd like to close by pointing out that the RNA-guided gene regulation we observe in CRISPR-Cas systems is fundamental to the way these systems work in bacteria, but also to the way they operate as technologies for genome editing and beyond.

Delivery and control are key to the future of genome editing. We need to have better ways to deliver CRISPR-Cas9 and related proteins into cell types of interest for genome editing, including into human patients. Fundamental research will continue to drive the field forward as it has in the past.

Figure. 23. The possibilities are endless.

I'm excited about what will happen in the future both with fundamental research, and with the applications that will solve real world problems in human health and the environment.

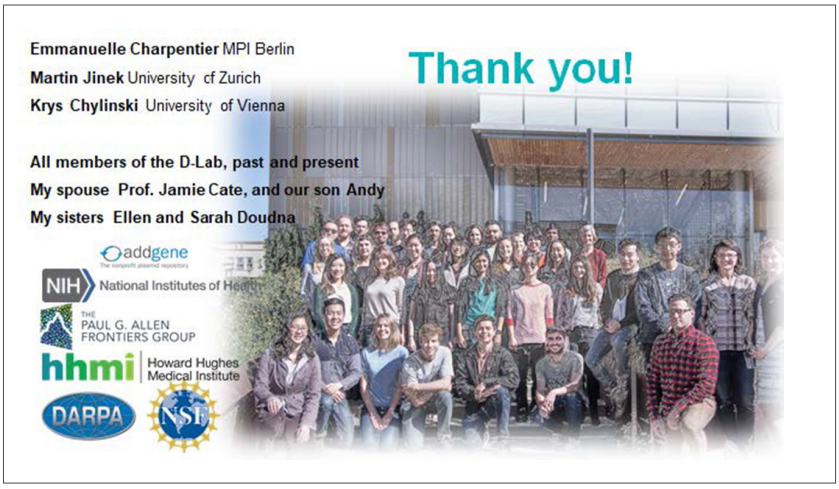

Figure. 24. Thanks.

I'd like to thank my past and current lab members with whom I've had the joy of doing science together over many years. I want to thank the many colleagues who have been involved in the fields of genome engineering, in DNA repair, and in the applications of genome editing that have made the field so exciting over the last few years. And finally, I want to thank my colleagues at the University of California, Berkeley, who I've had the joy of working with, and where we share together the dedication to public education and research that makes our work so rewarding. Thank you very much.